3GPP TSG-RAN WG4 Meeting #116bis R4-250

**Prague, Czech Republic, Oct. 13-17, 2025**

**Agenda item:** 6.11.1

**Source:** Moderator (Qualcomm)

**Title:** Topic summary for [116bis][112] R19 AI for air interface

**Document for:** Information

# Introduction

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

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

A WF summarizing the agreements from RAN4#116 was agreed in RP-2509098. The discussion will take into account these previous agreements.

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

## Companies’ contributions summary

|  |  |  |
| --- | --- | --- |
| **T-doc number** | **Company** | **Proposals / Observations** |
| [**R4-2513404**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_116bis/Docs/R4-2513404.zip) | Samsung | **Reporting delay for performance monitoring** **Proposal 1: RAN4 should define the reporting delay requirement for performance monitoring following the time line defined in RAN1 based on the configured inference report type and monitoring report type.****Reporting accuracy for performance monitoring****Observation 1: For performance monitoring metric report, SGCS will be used and is quantized with 4-bit. For each report, two SGCS are included to reflect the Quantity reported.****Proposal 2: RAN4 could consider to define the report delay requirement based on the ratio of SGCS reports during the whole testing time, and to check whether the ratio can be larger than pre-defined threshold Y with X% of test time, where the SGCS reporting is based on RAN1 definition.** **Mapping-table of reporting SGCS****Proposal 3: RAN4 should following mapping table deigned in RAN1 to define the reporting accuracy requirement** **Generalization** **Proposal 4: RAN4 should clarify how the generalization testing will be conducted, whether it will be tested in one case with multiple configurations combination, or it will be tested in multiple cases with individual configurations****Proposal 5: If one test case is conducted with multiple test configurations, one case per UE capability should be enough. If one test case is conducted with individual test configuration, multiple cases with different configurations can be considered.** **Proposal 6: If non-static scenario/configuration introduced, RAN4 should discuss how to perform the test with different parameters during one test setup.****Proposal 7: There is no need the generalization testing related with indoor/outdoor scenario.** **Proposal 8: There is no need the generalization testing related with channel bandwidth** **Observation 2: Rel-19 SCM channel model study is expected to only focus on limited scenarios. The testability and test repeatability for CDL channel model on demodulation and testing should be further studied****Proposal 9: Prioritize to use TDL channel model for AI/ML based CSI prediction for generalization testing in Rel-19. Further study the performance impact on AI/ML based 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 the specific channel model in Rel-20.****Proposal 10: RAN4 should generate the training data with the same gNB array parameters (NW-side) specified for test. For different gNB array parameters, RAN4 may need to define the multiple cases.****Observation 3: AI/ML based CSI prediction shows better performance compared with the nearest historical CSI scheme, while generalization problem related different doppler needs further investigated****Observation 4: With including the data from different doppler values for AI/ML model training, good generalization can be achieved for different doppler scenario, although there is a performance degradation** **Proposal 11: RAN4 can define requirements with considering different scenarios to guarantee the generalization of AL/ML based CSI prediction.****Proposal 12: RAN4 can consider to define different MCS requirement targeting different SNR range. The specific SNR range associated with MCS for generalization test should be further discussed.**  |
| [**R4-2513411**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_116bis/Docs/R4-2513411.zip) | vivo | **Proposal 1: For Type 3 performance monitoring for CSI prediction, RAN4 to discuss reporting accuracy requirements for SGCS 1 and SGCS 2.** * **For accuracy metrics for SGCS 1 and SGCS 2, extra non-prediction CSI report would be needed for TE to obtain the ground truth CSI, and CSI (non-predicted).**
* **SGCS 1 is calculated based on predicted CSI for one inference reporting, and ground truth CSI.**
* **SGCS 2 is based on ground truth CSI and CSI (non-predicted) corresponding to the latest CSI-RS transmission occasion not later than CSI reference resource of the inference reporting instance.**

**Proposal 2: For Type 3 performance monitoring for CSI prediction, RAN4 to define reporting delay requirements** **and measurement requirements, which could largely reuse the approach of Beam management monitoring.** |
| [**R4-2513420**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_116bis/Docs/R4-2513420.zip) | Huawei,HiSilicon | **Proposal 1**: CDL channel model is not used in RAN4 for AI CSI prediction.**Proposal 2:** For generalization testing, there is no need to use non-static condition.**Proposal 3:** RAN4 will start to discuss whether/how to define requirements for UE-side monitoring if other WGs achieve sufficient progress.**Proposal 4:** RAN4 will focus on NW-side monitoring for AI CSI prediction and legacy measurements reporting requirements are reused. |
| [**R4-2513539**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_116bis/Docs/R4-2513539.zip) | CAICT | ***Observation 1: To define the monitoring report delay, there are two possible alternatives,**** ***Alt 1:*** ***take the reception of triggering command as the stating point, and the transmission of first monitoring report as the ending point.***
* ***Alt 2:*** ***take the last RS occasion associated with corresponding monitoring report as the start point and the transmission of first monitoring report as the ending point.***

**Proposal 1: Suggest to discuss whether to consider monitoring measurement delay when defining monitoring report delay.*****Observation 2: The accuracy requirement of monitoring report including SGCS, could be reflected by measurement and prediction accuracy.*****Proposal 2: The definition and necessity of defining accuracy requirement for monitoring report requires clarification.** |
| [**R4-2513583**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_116bis/Docs/R4-2513583.zip) | Apple | 1. Dual-Metric Requirements for UE-Side CSI Prediction

**Ensure network trust in UE-reported SGCS and define LCM criteria by (i) enabling the test equipment (TE) to independently recompute SGCS1/SGCS2 and (ii) defining absolute and relative SGCS performance thresholds as triggers for LCM actions.*** **(Accuracy of reported KPI) Feasibility study of TE-based ground truth derivation: TE emulates known channels, derives SGCS1 and SGCS2 as reference values, and validates UE-reported results against this recomputed ground truth.**
* **(LCM) Absolute SGCS1 floor: Require that, over a defined monitoring window, a specified lower percentile of SGCS1 remains above an agreed minimum accuracy level. Scenario specific threshold**
* **(LCM )Dual-metric margin: Require that, within the same window, a high proportion of reports show SGCS1 exceeding SGCS2 by at least an agreed minimum margin.**
* **(LCM) Persistence and hysteresis: Trigger LCM fallback only when either requirement is violated consistently across consecutive monitoring windows and clear the condition after sustained compliance.**
* **(LCM) Per-layer and RI coupling: Apply checks per reported layer and adjust thresholds if the rank indicator changes.**
* **(LCM) Mobility/Doppler profiles: Provide profile-specific thresholds to reflect differences in channel dynamics.**
1. RAN4 should define a **reporting-delay requirement** as the elapsed time from completion of the RRC (re)configuration message, including AIML functionality activation and monitoring-resource setup, to the transmission of the first SGCS-based performance-monitoring report. This definition aligns with the beam-management precedent and captures all key delay contributors: configuration processing, model activation, and UE scheduling of the first monitoring instance.
 |
| [**R4-2513662**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_116bis/Docs/R4-2513662.zip) | CMCC | ***Proposal 1: for CSI prediction, RAN4 no need to define report mapping for SGCS, since RAN1 has defined the SGCS quantization mapping table in TS38.214.******Proposal 2: based on RAN1 agreements, for CSI prediction, the inference reporting delay is legacy Z/Z’ plus t, where t is reported by UE.***  |
| [**R4-2513765**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_116bis/Docs/R4-2513765.zip) | MediaTek inc. | **Observation #1: The performance gains of precoding prediction depend strongly on Doppler.****Proposal #1: Evaluate the performance‑monitoring metric under test conditions corresponding to the expected Doppler sweet spot (highest achievable gains).****Proposal #2: Evaluate the performance monitoring metric under test conditions with very low or very high Doppler, where no gains - or even a loss - are expected.** |
| [**R4-2513805**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_116bis/Docs/R4-2513805.zip) | OPPO | Proposal 1: For AI-based CSI prediction performance: when compared to R16 eTypeII using 'sample and hold' reporting, it should not be worse than R16 eTypeII.Proposal 2: Reuse the legacy PMI requirement (γ=1.8 for Enhanced Type II codebook for predicted PMI @ 20Hz Doppler spread), FFS γ value for 50Hz Doppler spread.**Minimum requirement for AI/ML based CSI prediction**

|  |  |  |
| --- | --- | --- |
| **Parameter** | **Test 1 (20Hz Doppler spread)** | **Test 2 (50Hz Doppler spread)** |
| $$γ\_{AI/ML}$$ | **1.8** | **TBD** |

Proposal 3: Within a CSI prediction window, $γ\_{AI/ML}$ is obtained by calculating $t\_{ue\\_AI/ML}$ and $t\_{rnd}$ within the CSI prediction window first and then get the averaged $γ\_{AI/ML}$ value.Proposal 4: 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 $γ\_{AI/ML}$ for performance [and calculates the measurement/inference latency for LCM]
 |
| [**R4-2513973**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_116bis/Docs/R4-2513973.zip) | Ericsson | Observations:[Observation 1: RAN1 has specified the SGCS quantization mapping table in TS 38.214 5.2.1.4.6.](#_Toc210418433)[Observation 2: The performance monitoring for CSI prediction (FG 58-3-5) is optional even if UE supports CSI prediction for UE-sided inference when N4=1 (FG 58-3-1).](#_Toc210418434)[Observation 3: If UE supports the CSI-PAI reporting, this UE should support AI/ML-based CSI prediction.](#_Toc210418435)[Observation 4: RAN1 has specified the measurement period for aperiodic/semi-persistence CSI-PAI reporting in TS 38.214.](#_Toc210418436)[Observation 5: It is not feasible to verify the measurement accuracy of SGCS1 and SGCS2 in TE.](#_Toc210418437)Proposals:[Proposal 1: RAN4 does not need to specify the SGCS quantization mapping table in TS 38.133.](#_Toc210418438)[Proposal 2: RAN4 only needs to refer to TS 38.214 5.2.1.4.6 if RAN4 will define the CSI-PAI reporting delay requirements in TS 38.133.](#_Toc210418439)[Proposal 3: Study the feasibility to set the CSI-PAI reporting requirements with that the SGCS1 should be more than X1 in [90]% of the test time with SNR=Y1 (dB), e.g., X1=0.85 and Y1=10dB.](#_Toc210418440)[Proposal 4: Study the feasibility to set the CSI-PAI reporting requirements with that the radio SGCS1/SGCS2 should be more than X2 in [90]% of the test time with SNR=Y2 (dB), e.g., X1=1.2 and Y1=10dB.](#_Toc210418441)[Proposal 5: If UE support CSI-PAI reporting capability, check the statistics of SGCS together with throughput ratio of follow predicted PMI and random Rel-15 TypeI PMI.](#_Toc210418442)[Proposal 6: RAN4 should consider at least two cases: one case for the AI/ML-based prediction works, and another case for the prediction does not work well.](#_Toc210418443) |
| [**R4-2514351**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_116bis/Docs/R4-2514351.zip) | Nokia | 1. RAN4 can define reporting accuracy requirements for AI/ML-based CSI prediction by using a test framework that monitors SGCS consistency under static radio conditions.

A potential test framework can be specified in RAN4, where the UE calculates and reports SGCS values under static and controlled radio conditions. Since the environment remains unchanged, any variation in reported values indicates potential inaccuracies in the UE’s reporting mechanism.1. RAN4 should define reporting accuracy requirements for CSI prediction performance monitoring using a test framework that evaluates the stability of reported values in a fixed environment.
 |
| [**R4-2514456**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_116bis/Docs/R4-2514456.zip) | Qualcomm Incorporated | **Observation 1:** RAN4 needs to finalize following aspects of CSI performance monitoring:* Selection of metric for requirement
* Reporting delay,
* Reporting accuracy, including the feasibility of defining reporting accuracy.
* Mapping table.

**Observation 2**: RAN1 agreed to define two SGCS values in performance monitoring. First SGCS is based on predicted CSI for one inference reporting and ground truth CSI. Second SGCS is based on ground truth CSI and non-predicted CSI corresponding to the latest CSI-RS occasion.**Observation 3:** The SGCS based on predicted CSI for one inference reporting and ground truth CSI conveys the performance of CSI prediction.**Observation 4:** The SGCS based on ground truth CSI and non-predicted CSI corresponding to the latest CSI-RS occasion conveys how well sample-and-hold based CSI feedback would perform.**Observation 5:** The ratio of the first and second SGCS of observation 2 conveys the ratio of the performance of CSI prediction to that of sample-and-hold to the network. **Observation 6:** The ratio of the first and second SGCS of observation 2 can become high even if first and second SGCS are low. This can happen specially when first and second SGCS are quite low; and a sudden fluctuation in one of these two metrics can make the ratio quite high.**Observation 7**: RAN1 has already defined the mapping table for reporting of performance monitoring. The RAN1 defined mapping table explicitly maps SGCS value, reported to convey the outcome of performance monitoring, to codepoints.**Observation 8:** Test equipment vendor cannot truly estimate the accuracy of UE’ reported SGCS during performance monitoring. Test equipment vendor can only check if the UE’s reported SGCS, averaged across many occasions, exceed a threshold where the threshold is defined based on simulations.**Observation 9:** Until Rel-18, RAN4 has primarily used TDL based channel models for demod testing, including requirements for PMI testing.**Observation 10:** RAN4 has selected TDL to define throughput ratio related requirements of AI-ML based CSI prediction.**Observation 11:** One of the main objectives of defining requirements for performance monitoring of CSI prediction is to inform gNB what it can expect in the field.**Proposal 1:** RAN4 selects the statistics of SGCS1, defined based on predicted CSI for inference reporting and ground truth CSI, as the metric to evaluate UE’s performance monitoring for CSI prediction. UE would pass a test if its reported SGCS1, averaged across many occasions, exceed a threshold.**Proposal 2:** RAN4 reuses the RAN1 defined mapping table for reporting of CSI prediction’s performance monitoring**Proposal 3:** RAN4 considers the following note while discussing to define the requirements for performance monitoring.* Note: Test equipment vendor cannot truly estimate the accuracy of UE’ reported SGCS during performance monitoring. Test equipment vendor can only check if the UE’s reported SGCS, averaged across many occasions, exceed a threshold where the threshold is defined based on simulations.

**Proposal 4:** RAN4 investigates UE’s reported SGCS1 during performance monitoring for CSI prediction across various channels, e.g., TDL-A, TDL-B and TDL-C, and selects the worst-case scenario to define the threshold for requirements. |
| [**R4-2513405**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_116bis/Docs/R4-2513405.zip) | Samsung | **Proposal 1: Existing test configuration with aperiodic CSI-RS configuration can be considered as starting point for AI/ML based CSI prediction.****Proposal 2: Focus on periodic CSI-RS configuration for AI/ML based CSI prediction firstly, after the alignment work of periodic CSI-RS Configuration is stable, RAN4 can proceed the related alignment and requirement discussion for Aperiodic CSI-RS configuration** **Proposal 3: Considering only 3 meeting is left based on the original plan, if TDD requirement with AI/ML based on CSI prediction also need to be specified in Rel-19, it is better to discuss the related test configuration and prepare the AI/ML model alignment.****Proposal 4: Pending on the results, companies can discuss the MCS selected for requirement if it is feasible for testing.****Proposal 5: RAN4 will update the schedule timeline for requirement according to the value of t pending on UE capability report****Observation 1: With mix dataset training, the performance has minor performance degradation. Good generalization performance can be achieved with mix dataset training****Observation 2: Compared with Rel-15 Type I random PMI, both AI and non-AI schemes can achieve large performance gain for different doppler scenario. Compared with non-AI scheme Rel-18 eType II Doppler with K=1, the gain of AI scheme is minor, especially under the scenario with doppler as 50Hz** **Observation 3: With doppler increasing, all the schemes show the performance degradation** **Observation 4: With the AI/ML model training under high SNR condition, better performance can be achieved compared with the AI/ML model training with the test SNR****Proposal 6: RAN4 should further check the generalization related with SNR for AI/ML based CSI prediction****Proposal 7: RAN4 should clarify the SNR range of dataset for AI/ML model training if the final results among companies have large gap.**  |
| [**R4-2513416**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_116bis/Docs/R4-2513416.zip) | vivo | **Observation 1: Initial evaluation results for CSI prediction AI/ML model is shown in Table 2-1, for FDD and 32gNB Tx.** Table 2-1. Initial evaluation results for CSI prediction AI/ML model, for FDD and 32gNB Tx.

|  |  |  |  |
| --- | --- | --- | --- |
| FDD, 32 gNB Tx | Test on 20Hz Doppler spread | Test on 50Hz Doppler spread | Test on 100Hz Doppler spread |
| Model trained by 20Hz Doppler spread | 0.993 | / | / |
| Model trained by 50Hz Doppler spread | / | 0.902 | / |
| Model trained by 100Hz Doppler spread | / | / | 0.298 |
| Model trained by 20Hz, 50Hz and 100Hz Doppler spread | 0.994 | 0.907 | 0.296 |

**Proposal 1: Reference AI model for CSI prediction would be needed for performance requirement definition, if it is difficult to align the simulation results.** |
| [**R4-2513424**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_116bis/Docs/R4-2513424.zip) | Huawei,HiSilicon | **Proposal 1**: For step-1, SGCS performance results of AI CSI prediction are provided in Table 1.Table 1. SGCS performance results of AI CSI prediction (step-1)

|  |  |  |
| --- | --- | --- |
|  | 20Hz | 50Hz |
| Rank 1 | 0.9985413 | 0.8673333 |
| Rank 2 | 0.99800384 | 0.8166052 |

 |
| [**R4-2513587**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_116bis/Docs/R4-2513587.zip) | Apple | 1. Curriculum-Based SNR Training for AI/ML CSI Prediction

**RAN4 is invited to perform CSI-prediction evaluation by combining clear test assumptions with a staged SNR training approach that improves model robustness to noise*** **Baseline test conditions – Standardize practical channel estimation and SNR operating ranges to define the conditions under which performance requirements (e.g., SGCS thresholds, throughput ratios) will later be applied.**
* **Curriculum SNR training – Begin with high-SNR samples where CSI structure is clean, then progressively expand to moderate and lower SNR ranges to strengthen robustness.**
* **Improved convergence and generalization – Enable the model to learn stable spatial-temporal patterns first, reducing overfitting to noise and improving performance under realistic conditions.**
1. Study Generalization of AI/ML CSI Prediction Across Deployment Variations

**RAN4 is invited to evaluate how well AI/ML CSI prediction models generalize when training and deployment conditions differ, as this determines the scope of performance verification.*** **Doppler spread / UE mobility – Examine robustness of prediction under varying mobility levels and channel dynamics.**
* **SNR range**
* **Impact on testing – Use findings to decide how many and which test cases are needed, allowing a more efficient but still reliable conformance framework**.
 |
| [**R4-2513663**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_116bis/Docs/R4-2513663.zip) | CMCC | ***Proposal 1: for CSI prediction test with static condition, it is proposed to define test cases with different configuration/parameters, e.g. antenna configuration, observation window length, prediction window length, and UE speed .*** ***Proposal 2: For channel model for CSI prediction tests, it is proposed to focus on TDL channel models in Rel-19. Using CDL channel model for CSI prediction test can be further discussed in later release based on the progress on SCM.***  |
| [**R4-2513766**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_116bis/Docs/R4-2513766.zip) | MediaTek inc. | **Observation #1: FDD 16TX and FDD 32TX results with the same Doppler perform quite similarly.****Observation #2: In lower Doppler accuracy is better than in higher Doppler.****Observation #3: Rel-18 codebook quantization performance is close to raw channel prediction performance.****Observation #4: In Step‑3, channel prediction shows gains over sample and hold in all tests when evaluated using SGCS.****Observation #5: In Step‑3, the SGCS measured gains of channel prediction over sample and hold are larger at 50 Hz than at 20 Hz.****Observation #6: Current RAN1 specification TS38.214 defines only a slot interval for which PMI is predicted while exact prediction slot is undefined.****Observation #7: For the test case, current RAN1 specification TS38.214 does not allow signalling the same five-slot period for PMI reporting that is used for applying the reported PMI at TX.****Observation #8: Prediction gain measured in SNR varies between 1.8 to 2.1dB in Doppler 20Hz.****Observation #9: Prediction gain measured in γ1 varies between 2.6 to 3.1 in Doppler 20Hz.****Observation #10: There is no prediction gain in Doppler 50Hz.****Observation #11: Operation point of MCS13 in Doppler 20Hz is close to SNR 0dB.****Observation #12: Optimized prediction timing can improve performance approximately 1dB.****Proposal #1: We propose to introduce PMI reporting requirement with Doppler 20Hz.****Proposal #2: We propose not to introduce PMI reporting requirement with Doppler 50Hz.****Proposal #3: We propose to use MCS17 or MCS19 for robust testing SNR.** |
| [**R4-2513807**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_116bis/Docs/R4-2513807.zip) | OPPO | **Observation 1: Regarding the updated simulation for AI/ML-based CSI prediction, the updated simulation results are shown in Table 5 for MCS=17 and Table 6 for MCS=19.****Observation 2: For step-2, the SGCS calculated by comparing the feedbacked SVD of the model output with the SVD of the ground truth raw channel deteriorates significantly. The SGCS calculated by comparing the feedbacked SVD of the model output with the feedbacked SVD of the ground truth raw channel shows much less degradation.****Observation 3: For step-3, under the condition of TX=16 and rank=2, the throughput performance obtained by randomly selecting a Type I codebook is relatively poor. The key difference between these schemes (e.g. model from different companies) will primarily be reflected in their γ values relative to the baseline benchmark.** |
| [**R4-2513974**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_116bis/Docs/R4-2513974.zip) | Ericsson | Observations:[Observation 1: Throughput ratio of AI/ML-based PMI prediction over Rel-15 TypeI PMI (γ1\_AI/ML) is larger for 16Tx compared with 32Tx.](#_Toc210419376)[Observation 2: Throughput ratio of AI/ML-based PMI prediction over Rel-16 eTypeII PMI (γ2\_AI/ML) is larger for higher MCS with TDLA30-20.](#_Toc210419377)[Observation 3: Higer SGCS does not always mean higher throughput with predicted PMI compared with Rel-16 eTypeII.](#_Toc210419378)[Observation 4: RAN1 is still discussing the CSI reporting delay for Rel-19 AI/ML-based PMI prediction.](#_Toc210419379)Proposals:[Proposal 1: RAN4 should define AI/ML-based PMI prediction performance requirements at least with the following configuration: 16 CSI-RS ports, TDLA30-20, MCS19 or MCS22.](#_Toc210419380)[Proposal 2: RAN4 evaluates other scenarios such as TDLC300-20/50 and/or other SNR test points to ensure the model generalization.](#_Toc210419381)[Proposal 3: RAN4 should evaluate the AI/ML-based PMI prediction performance based on the CSI processing time with t=0. Once RAN1 concludes the UE capability of reporting delay t, RAN4 should revisit the scheduling of CSI-RS, PMI reporting, and PDSCH.](#_Toc210419382) |
| [**R4-2514460**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_116bis/Docs/R4-2514460.zip) | Qualcomm Incorporated | **Observation 1: For 25Hz and 50Hz Doppler frequencies, the SGCS obtained with AI-ML based CSI prediction in both step 1 and step 2 are significantly higher than those obtained with “sample and hold” approach.****Observation 2: The SGCS performances of AI-ML model, trained based on the mixed dataset of 20, 50 and 100Hz, are almost identical to those of AI-ML model, trained based on individual datasets.****Observation 3: The number of parameters of our AI-ML CSI prediction model, that got used to generate the results of this contribution, is in the order of several millions.**  |

## Open issues summary

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

Agreements in R4-2511890 regarding CSI reporting framework and CSI prediction:

**Issue 1-5: Scheduling delay**

Agreement:

* + The delay from when the UE prediction is send until the time the TE applies is n+4 (slots) for FDD

**Issue 1-1: Performance monitoring**

Agreement:

RAN4 will introduce requirements for Type 3 performance monitoring for CSI prediction

FFS which requirements are to be introduced.

**Issue 1-2: Requirement baseline for monitoring**

Agreement:

Introduce the following requirements for CSI performance monitoring:

* Reporting delay
	+ FFS how the delay is defined
* Reporting accuracy
	+ FFS on whether accuracy requirement can be defined/checked
* Mapping table – To be checked whether RAN1 captures or RAN4
* If it is not feasible to define reporting accuracy requirement or test it, RAN4 will send an LS to RAN1 to inform RAN1 about this

**Issue 1-4: Reporting delay requirement**

Agreement:

wait for RAN1 decision

Also, simulation assumptions were agreed in R4-25xxxxx.

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

1. Requirements for performance monitoring
2. Reporting delay
3. TS for performance monitoring
4. Core requirement for PMI reporting
5. Performance monitoring accuracy metric
6. Simulation results
7. Doppler and SNR/MCS
8. Generalization, different SNRs
9. Activation

### Sub-topic 1-1

*Requirements for performance monitoring*

In the previous meeting it was agreed to introduce requirements for performance monitoring, however, the details are still up for discussion.

Below are some tables summarizing the relationship between the monitoring report type, inference report type and monitoring measurement type.

|  |  |  |
| --- | --- | --- |
| Monitoring report typeInference report type | SP report | AP report |
| AP report | Not support | Support  |
| SP report | Support | Support |

|  |  |  |
| --- | --- | --- |
| Monitoring report typeMonitoring measurement type | SP report | AP report |
| Periodic | Support | Support |
| Semi-persistent | Support | Support |
| Aperiodic | Not support | Support |

**Issue 1-1: Requirements for performance monitoring**

* Proposals
	+ Option 1: Introduce requirements for both semi-periodic and aperiodic reporting
	+ Option 2: Others
* Recommended WF

Option 1

### Sub-topic 1-2

*Reporting delay requirements for performance monitoring*

There are multiple proposals for the reporting delay requirements. Besides the options listed below, it would also need to be discussed whether there is a need to differentiate between AP and SP reporting.

**Issue 1-2: Reporting delay requirements for performance monitoring**

* Proposals
	+ Option 1: Follow the RAN1 defined timeline based on the configured inference report type and monitoring report type
	+ Option 2: take the reception of triggering command as the stating point, and the transmission of first monitoring report as the ending point
	+ Option 3: take the last RS occasion associated with corresponding monitoring report as the start point and the transmission of first monitoring report as the ending point
	+ Option 4: elapsed time from completion of the RRC (re)configuration message, including AIML functionality activation and monitoring-resource setup, to the transmission of the first SGCS-based performance-monitoring report.
	+ Option 5: others
* Recommended WF
	+ Option 3

### Sub-topic 1-3

*TS for performance monitoring*

The TS in which the CSI performance monitoring will be capture have to be agreed

**Issue 1-3: TS for performance monitoring**

* Proposals
	+ Option 1: 38.101-4
	+ Option 2: 38.133
		- 2a: introduce a new Clause with “Performance monitoring for AI/ML features/use cases”
		- 2b: under an existing clause, TBD
	+ Option 3: others
* Recommended WF
	+ To be discussed

### Sub-topic 1-4

*Core requirements for PMI prediction reporting*

The PMI reporting framework/test metric is a core requirement which should be finalized as soon as possible. The general Reporting of PMI is defined in 38.101-4 Clause 6.3. The test metric definition (γ)was already agreed, however this will have to be included in the specifications.

**Issue 1-4: Core requirements for PMI prediction reporting**

* Proposals
	+ Option 1: Introduce a new clause under Clause 6 (6.X) in 38.101-4 as “Reporting of predicted PMI”
	+ Option 2: Introduce a new subclause under 6.3 (6.3.X) in 38.101-4 as “Reporting of predicted PMI”
	+ Option 3: Introduce a new Clause 12 as “CSI prediction reporting requirements” and 12.X as “Reporting of predicted PMI” in 38.101-4
	+ Option 4: others
* Recommended WF
	+ To be discussed

The requirements will have to be drafted based on the agreed definition of γ in previous meetings.

Company to provide the CR in RAN#117: ??

### Sub-topic 1-5

*Performance monitoring accuracy metric*

Several companies proposed different ways to introduce requirements and corresponding tests for the performance monitoring accuracy.

**Issue 1-5: Performance monitoring accuracy metric**

* Proposals
	+ Option 1: RAN4 could consider defining the report delay requirement based on the ratio of SGCS reports during the whole testing time, and to check whether the ratio can be larger than pre-defined threshold Y with X% of test time, where the SGCS reporting is based on RAN1 definition.
	+ Option 2: Introduce accuracy requirements for SGCS1 and SGCS2 separately
		- SGCS 1 is calculated based on predicted CSI for one inference reporting, and ground truth CSI.
		- SGCS 2 is based on ground truth CSI and CSI (non-predicted) corresponding to the latest CSI-RS transmission occasion not later than CSI reference resource of the inference reporting instance.
	+ Option 3: Ensure network trust in UE-reported SGCS and define LCM criteria by (i) enabling the test equipment (TE) to independently recompute SGCS1/SGCS2 and (ii) defining absolute and relative SGCS performance thresholds as triggers for LCM actions.
	+ Option 4: Study the feasibility to set the CSI-PAI reporting requirements with that the SGCS1 should be more than X1 in [90]% of the test time with SNR=Y1 (dB), e.g., X1=0.85 and Y1=10dB.
		- Study the feasibility to set the CSI-PAI reporting requirements with that the radio SGCS1/SGCS2 should be more than X2 in [90]% of the test time with SNR=Y2 (dB), e.g., X1=1.2 and Y1=10dB.
	+ Option 5: RAN4 should define reporting accuracy requirements for CSI prediction performance monitoring using a test framework that evaluates the stability of reported values in a fixed environment.
	+ Option 6: RAN4 selects the statistics of SGCS1, defined based on predicted CSI for inference reporting and ground truth CSI, as the metric to evaluate UE’s performance monitoring for CSI prediction. UE would pass a test if its reported SGCS1, averaged across many occasions, exceed a threshold.
		- RAN4 investigates UE’s reported SGCS1 during performance monitoring for CSI prediction across various channels, e.g., TDL-A, TDL-B and TDL-C, and selects the worst-case scenario to define the threshold for requirements.
	+ Option 7: other proposals
* Recommended WF
	+ To be discussed

### Sub-topic 1-6

*Simulation results and next steps*

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

**Issue 1-6: Simulation results and next stepts**

* Proposals
	+ Discuss the simulation results and next steps
		- Simulation results
		- Refinement of parameters:
			* Introduce realistic channel estimation
			* Further discuss CSI-RS configuration, codebook configuration, Dopller, SNR, etc
* Recommended WF
	+ To be discussed

Discussion on refinement of simulation parameters to be done mainly offline

### Sub-topic 1-7

*Doppler and MCS choiceGeneralization*

**Issue 1-7: Doppler and MCS**

* Proposals
	+ Option 1: Introduce tests for 20Hz Doppler with MCS17/19 with 32Tx ports
	+ Option 2: Introduce tests for 50Hz Doppler with MCS17/19 with 32Tx ports
	+ Option 3: Others
* Recommended WF
	+ Option 1

### Sub-topic 1-8

*Generalization*

Some companies have proposed introduction of tests at different SNR levels or vary other parameters(e.g. Doppler) to ensure robustness of the UE model.

**Issue 1-8: Generalization**

* Proposals
	+ Option 1: Introduce tests targeting different SNR points
	+ Option 2: Introduce tests with different channel models (e.g. TDL-C 300-20/50)
	+ Option 3: others
* Recommended WF
	+ To be discussed

The need for generalization tests and what should be discussed/studied should be debated.

### Sub-topic 1-9

*Activation*

Some companies have proposed the introduction activation requirements.

**Issue 1-9: Activation**

* Proposals
	+ Option 1: Introduce activation requirement, agreements on activation for beam prediction is applied for CSI prediction. In detail, for CSI prediction, UE has to be ready to start measurements for inference after sending RRC reconfiguration complete.
	+ Option 2: No need for anything new, we currently do not have activation for legacy CSI reporting
	+ Option 3: others
* Recommended WF
	+ To be discussed

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

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

## Companies’ contributions summary

|  |  |  |
| --- | --- | --- |
| **T-doc number** | **Company** | **Proposals / Observations** |
| [**R4-2513080**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_116bis/Docs/R4-2513080.zip) | MediaTek Inc. | **Proposal 1**: For AI/ML BM, inference delay in the is Z3’+d’ as agreed and to-be-defined in RAN1.**Proposal 2**: For AI/ML BM case 2, the length of observation window is up to UE implementation. No need to define the delay for the first valid prediction report for periodic and semi-persistent report.**Proposal 3**: In the observation period agreed last meeting, T = 1 for AI/ML BM case 2. No need to have a new chapter to capture the requirement of observation period as it is the same as legacy measurement period defined in 9.5.4 in 38.133.**Proposal 4**: If the predicted Tx beam in Set A is QCL Type-D to a known Tx beam, then legacy known TCI state conditions are applicable. **Proposal 5**: Use a separate paragraph to capture the known TCI state conditions for the case “the predicted Tx beam in Set A is not QCL Type-D to a known Tx beam”. For this case, known TCI state conditions can be defined as- UE reports [TCI state known] in Capability X, and- TCI state switch command is received within 1280 ms upon transmitting the predicted L1-RSRP report for the target TCI state- The TCI state remains detectable during the TCI state switching period, and- The SSB associated with the TCI state remain detectable during the TCI switching period- SNR of the TCI state ≥ -3 dB**Proposal 6**: Both BB error and RF error should be considered when define the requirements and test case.**Observation 1**: The prediction accuracy is improved a bit by increasing the training dataset size.**Observation 2**: Using different dataset, for some metric, the accuracy performance are similar, while for some other metric, the accuracy performance may have some divergence.**Proposal 7**: Prefer that the metric to be used in RAN4 is also used for monitoring, i.e., x shall be 0.**Proposal 8**: RSRP accuracy requirements of Top-1 of predicted beams can be defined by filtering out the data with SNR of the Top-1 beam in set A no less than -3dB. **Observation 3**: For different beams, e.g., predicted 1st strongest and 5th strongest beam, the predicted absolute RSRP accuracy varies. It may be the SINR difference that results in the difference between the predicted absolute RSRP accuracy of top-1 beam and other beams.**Proposal 9**: Further discuss whether it is workable to define a single absolute accuracy requirement which is applicable to all/some of the predicted beams with some side conditions.**Observation 4**: The report result for model monitoring is to reflect whether the model in use works well in the current scenario. **Proposal 10**: It is not possible to define the accuracy requirements for model monitoring for AI/ML BM.**Proposal 11**: It is not necessary to define test cases for model monitoring for AI/ML BM, as no UE will fail the test.**Observation 5**: CDL channel includes two parts “from BS to clusters” and the channel “from clusters to UE”.**Proposal 12**: Discuss how to simulate the channel “from BS to clusters” in the simplified CDL-channel as well.**Proposal 13**: Discuss how to emulate UE movement in multiple AoA test systems.**Observation 6**: It is not workable to emulate UE movement through rotating UE during the test as the time needed to rotate UE is about 1s.**Proposal 14**: Evaluate whether it is workable to emulate UE movement through adjusting TE transmission power at each probe in multiple AoA test systems.**Proposal 15**: The procedures to emulate UE movement using simplified CDL channel in the simulation will be:* Cast the UE in the cell randomly.
* Generate the channel between BS and UE according to 38.901, including path-loss, shadowing, and small-scale fading.
* Simplify small scale fading between BS and UE based on the to-be-agreed simplification method.
* Calculate received RSRP of each beam.

**Proposal 16**: Discuss how to get training dataset if to use multiple AoA test systems:* Alt 1: collect the training data in the chamber for each test UE before training
* Alt 2: generate the training dataset assuming a certain UE radiation pattern through simulation

**Observation 7**: Multiple AoA test systems cannot use peak fine beam direction only.**Observation 8**: The upper bound of SNR with multiple AoA test system is 13dB. **Proposal 17**: Take the upper bound of SNR with multiple AoA test system into consideration during simulation.  |
| [**R4-2513217**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_116bis/Docs/R4-2513217.zip) | CATT | **Proposal 1: For the case where both predicted beam ID(s) and predicted RSRP(s) are reported, the related tests apply to the top-K beams, where K is the valued configured by *nrofreportedpredictedrs-r19*.** **Proposal 2: RAN4 to verify at least half of predicted beams, i.e., N=**$ \left⌊^{K}/\_{2}\right⌋+1$**.** **Proposal 3: The same x value applies for each comparison between the same ranked predicted beam and genie-aided beam. Value of x is further discussed based on the simulation results.** **Proposal 4: RAN4 not to apply absolute RSRP accuracy requirement to other beams, i.e., predicted beams with lower RSRPs.** **Proposal 5: RAN4 to remove the square bracket in the definition of relative RSRP accuracy, i.e.,****Relative RSRP accuracy for reported beams during inference reporting = (predicted L1-RSRP of beam index i - reported L1-RSRP of beam index n) - (ground truth of L1-RSRP of beam index i - ground truth of L1-RSRP of beam index n), ~~[~~where the beam index n owns the largest reported value~~]~~.****Proposal 6: The** **reported L1-RSRP in the relative RSRP accuracy definition can be measured RSRP.** **Proposal 7: The relative RSRP accuracy definition applies for BM-Case 2.** **Observation 1: Different values for T may be needed depending on the value of M and prediction configurations.** **Proposal 8: RAN4 to specify the value(s) of T considering the following alternatives:*** **Alt 1: Specify a fixed product of T\*M;**
* **Alt 2: Specify two different values for M=1 and M=3 respectively;**

**RAN4 deprioritizes the usage of UE capability for defining T.****Proposal 9: RAN4 to define a fixed value for the time duration of inference delay procedure.**  |
| [**R4-2513412**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_116bis/Docs/R4-2513412.zip) | vivo | **Proposal 1：For prediction reporting delay requirement for BM-Case 2, it is not applicable for aperiodic CSI-RS reporting scenario.****Proposal 2: RAN4 to define unified Prediction report delay, which applies to semi-persistent,periodic and periodic reports, and it can be defined as*** **For CSI-RS based prediction , the prediction period requirements TL1-RSRP\_Prediction\_Period\_CSI-RS for BM can be defined as**

Table 1: Prediction period TL1-RSRP\_Prediction\_Period\_CSI-RS for FR2

|  |  |
| --- | --- |
| Configuration | TL1-RSRP\_Prediction\_Period\_CSI-RS (ms)  |
| non-DRX | max(TReport, ceil(K\*M\*P\*N)\*TCSI-RS+Tinference)  |
| DRX cycle ≤ 320 ms | max(TReport, ceil(1.5\*K\*M\*P\*N)\*max(TDRX,TCSI-RS) +Tinference) |
| DRX cycle > 320 ms | ceil(K\*M\*P\*N)\*TDRX +Tinference |
| NOTE 1: TCSI-RS is the periodicity of CSI-RS configured for L1-RSRP measurement. TDRX is the DRX cycle length. TReport is configured periodicity for reporting. Tinference is the time required for model inference, which is defined by d/d' as specified according to clause 5.4 in TS 38.214 [26].NOTE 2: the requirements are applicable provided that the CSI-RS resource configured for L1-RSRP measurement is transmitted with Density = 3. |

* **For SSB-based prediction, the prediction period requirements TL1-RSRP\_Prediction\_Period\_SSB for BM can be defined as**

Table 2: Prediction period TL1-RSRP\_Prediction\_Period\_SSB for FR2

|  |  |
| --- | --- |
| Configuration | TL1-RSRP\_Prediction\_Period\_SSB (ms)  |
| non-DRX | Max (TReport, ceil (K\*M\*P\*N) \*TSSB+Tinference) |
| DRX cycle ≤ 320 ms | Max (TReport, ceil (1.5\* K\*M\*P\*N) \*max (TDRX, TSSB) +Tinference) |
| DRX cycle > 320 ms | Ceil (1.5\* K\*M\*P\*N) \*TDRX+Tinference |
| Note: TSSB = *ssb-periodicityServingCell* is the periodicity of the SSB-Index configured for L1-RSRP measurement. TDRX is the DRX cycle length. TReport is configured periodicity for reporting. Tinference is the time required for model inference, which is defined by d/d' as specified according to clause 5.4 in TS 38.214 [26]. |

**Where,for BM-Case 1, K=1; for BM-Case 2, the value of K depends on the UE capability on the number of latest consecutive transmission occasions designed by RAN1.****Observation 1: In RAN1 design, for a CSI-ReportConfig with reportQuantity-r19, it can be set to 'p-ssb-index-r19' or 'p-ssb-index-RSRP-r19', which means it supports the scenario of predicting SSB.** **Proposal 3: RAN4 needs to clarify whether the requirement for the core part (e.g., in the prediction delay requirement) considers the scenario of predicting SSB.****Proposal 4: For the agreed time conditions for TCI state, further clarification on ‘the last transmission of the RS resource’ is as follows:*** **TCI state switch command is received within 1280 ms upon the last transmission of the RS resource, where the RS resource here refers to**
* **the RS resource used for the L1-RSRP measurement reporting for the target TCI state [legacy case] or**
* **the RS resource in *resourcesForChannelMeasurement* used for prediction of *resourcesForChannelPrediction-r19* [The RS resource for set B measurements]**

**Proposal 5: RAN4 to update the known TCI state conditions as** *The TCI state is known if the following conditions are met:**- During the period from T0 ~~the last transmission of the RS resource used for the L1-RSRP measurement reporting for the target TCI state~~ 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~~**- TCI state switch command is received within 1280 ms upon ~~the last transmission of the RS resource for beam reporting or measurement~~ T0**- The UE has sent at least 1 L1-RSRP measurement or predicted RS quantities report for the target TCI state 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 ≥ -3 dB**Where T0 is**- the last transmission of the RS resource used for the L1-RSRP measurement reporting for the target TCI state, where the RS resource for L1-RSRP measurement is the RS in target TCI state or QCLed to the target TCI state [legacy case]**- or the last transmission of the RS resource in resourcesForChannelMeasurement used for prediction of resourcesForChannelPrediction-r19, where**- the RS resource in resourcesForChannelMeasurement used for prediction of resourcesForChannelPrediction-r19 is the RS in target TCI state or QCLed to the target TCI state or**- the one of RS resource in resourcesForChannelPrediction-r19 is the RS in the target TCI state or QCLed to the target TCI state but not in the RS resource in resourcesForChannelMeasurement, and UE supports [known TCI based on prediction capability],**Otherwise, the TCI state is unknown.***Proposal 6: RAN4 to define absolute RSRP accuracy as one of metrics for beam management requirements/tests, the corresponding definition is:*** ***For BM-Case 1, the absolute RSRP accuracy = predicted L1-RSRP of beam index i – ground-truth of L1-RSRP of beam index i.***
* ***For BM-Case 2, the absolute RSRP accuracy = predicted L1-RSRP of beam index i for time instance m – ground-truth of L1-RSRP of beam index i for time instance m***
* ***Where, The index i corresponds to the index of Top-1 beam of predicted beams; the index m is corresponds to the index of time instance, 1<=m<=M***

**Proposal 7: RAN4 to use the follow definition to define the relative RSRP accuracy requirement:** * ***For BM-Case 1, Relative RSRP accuracy for reported beams during inference reporting = (predicted L1-RSRP of beam index i - Reported L1-RSRP of beam index n) - (ground truth of L1-RSRP of beam index i - ground truth of L1-RSRP of beam index n)***
* ***For BM-Case 2, the absolute RSRP accuracy = predicted L1-RSRP of beam index i for time instance m – ground-truth of L1-RSRP of beam index i for time instance m.***
* ***Where, The index i corresponds to the index of Top-K (K is the number of reported predicted beam) beam of predicted beams and i >1; The beam index n owns the largest reported value; The index m is corresponds to the index of time instance, 1<=m<=M***
* ***The reported L1-RSRP refers to predicted L1-RSRP***

**Proposal 8: For Metrics/KPIs for beam ID prediction agreed in RAN4#114bis, RAN4 not to further consider the verification of top-N (N</=K) predicted beams in the success rate evaluation****Proposal 9: For Metrics/KPIs for beam ID prediction agreed in RAN4#114bis, it will apply to the case only beam ID is reported and the case both beam ID and L1-RSRP are reported.****Proposal 10: For the X value of Metrics/KPIs for beam ID prediction agreed in RAN4#114bis, it can be determined based on the simulation results** |
| [**R4-2513413**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_116bis/Docs/R4-2513413.zip) | vivo | **Draft CR** |
| [**R4-2513421**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_116bis/Docs/R4-2513421.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, take the existing core requirement for beam related information reporting as the starting point.**Proposal 3:** RAN4 will focus on NW-side performance monitoring Type 1 option 1 and reuse legacy requirements for measurement reporting. |
| [**R4-2513453**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_116bis/Docs/R4-2513453.zip) | Xiaomi | **Observation 1: It is already defined in 38.214 that the number of measurement instances (K\_BM) required for a beam management report is dependent on UE capability.****Proposal 1: For BM-case 2, introduce a scaling factor Q to adjust the measurement time for RSRP predictions, where Q is determined by UE capability, to adopt to different UE implementations.****Proposal 2: For simplified CDL model, cluster number can be reduced from spatial angle offset and power aspect:*** **Combine clusters if AOA offset between these clusters is smaller than a threshold**
* **Remove cluster with low power.**

**Observation 2: For simplified CDL-C model, L1-RSRP of weaker TX beams will decrease more than strong TX beams when compared with CDL-C model. The impact on AI prediction needs further study.****Observation 3: For simplified CDL-C model, L1-RSRP delta between different TX beams will increase if cluster number and AOA number is further reduced. It's easier for AI to distinguish the best Top-K TX beam. The AI prediction requirement will be relaxed.****Proposal 3: For simplified CDL model, RAN4 to choose 8 clusters and 3 AOA numbers to balance test complexity and prediction performance requirement.****Proposal 4: If the test metric is beam index prediction accuracy, UE may need to report ground truth of N best beam index for set A.****Observation 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 5: 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 6: For BM case-1, channel doppler can set to 0 or a small value to guarantee that there is neglectable L1-RSRP variation.****Proposal 7: 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 8: RAN4 to discuss how to solve the UE cheating issue if UE report both predicted result and ground truth.** |
| [**R4-2513584**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_116bis/Docs/R4-2513584.zip) | Apple | 1. RAN4 shall update the definition of beam prediction reporting timelines to explicitly include the inference delay introduced by AI/ML models. This delay shall be aligned with the parameter d’ defined in RAN1, which extends the legacy Z3’ timing between reference signals and L1-RSRP reporting. By referencing d’ directly, RAN4 ensures consistency with RAN1 specifications, avoids duplicate definitions, and provides a clear basis for conformance testing
2. RAN4 should define LCM requirements for UE-assisted monitoring where the UE reports the RS-PAI metric (monitoring Type 1 Option 2) in BM-Case 1 and BM-Case 2. The UE-reported N\_p/N ratio must exceed a minimum threshold.
3. RAN4 to define requirements for **end-to-end RS-PAI reporting delay**: start when the UE either (a) receives the configuration of monitoring RS resources or (b) sends RRC Reconfiguration Complete, and end when the first RS-PAI report is received at the gNB. This delay must bound the total time for window accumulation, UE processing, and uplink scheduling, giving the network a deterministic basis for resource planning and timely fallback.
4. RAN4 Requirement on LCM Testing and Trigger Criteria Using RS-PAI

**RAN4 shall include Life-Cycle Management (LCM) testing in its performance requirements for AI/ML beam-management models.*** **Primary KPI**: **Use RS-PAI (N\_p / N) as the single KPI to monitor live prediction accuracy during testing.**
* **Trigger thresholds**: **Define minimum RS-PAI ratios and observation windows (e.g., sustained below X % over Y consecutive reports) that, when breached, trigger an LCM action.**
* **LCM actions to test**: **Model switching, adapter activation/re-training, or deactivation with fallback to legacy beam management.**
* **Extended conditions**: **Include criteria for persistent instability (e.g., large RS-PAI variance) and deployment changes (e.g., new SSB/CSI-RS configuration) to test the ability to pre-emptively trigger LCM.**

**This proposal ensures that RAN4 tests not only accuracy but the full life-cycle response—verifying that RS-PAI based thresholds reliably activate switching, deactivation, or fallback in conformance testing and field operation.**1. RAN4 Requirement on LCM Delay for Model Switching, Activation and DeactivationRAN4 shall specify end-to-end delay requirements for LCM actions in AI/ML beam-management.
* **LCM actions covered**: **model switching, new-model activation, and model deactivation with fallback to legacy operation.**
* **Start and end points**: **The delay measurement shall start when the LCM trigger condition (e.g., RS-PAI threshold breach or deployment-change event) is detected by the gNB or UE and shall end when the target model is fully active or legacy fallback is completed.**
* **Maximum delay bounds**: **RAN4 shall define maximum acceptable delays (e.g., X ms for deactivation/fallback, Y ms for model/functionality switching) to guarantee service continuity.**
* **Test method**: **LCM delay shall be validated in conformance testing under both static and dynamic radio conditions, including the time for signaling, UE processing, model load, and uplink scheduling.**
1. Two-Phase RAN4 Conformance Test for AI/ML Beam Management: RAN4 should adopt a **two-phase test** to validate AI/ML beam-prediction accuracy and full UE behaviour for BM-Case 1 (spatial) and BM-Case 2 (temporal).
* **Phase 1 – Algorithmic validation (single-AoA “wireless-cable” setup)**
	+ **Fading emulator with full CDL-A/B/C channels for rich, spatially consistent mobility.**
	+ **Tests algorithm only, independent of UE beam-sweeping hardware.**
	+ **Verifies train/test channel combinations (full↔full, simplified↔simplified, full↔simplified) and all AI input modes.**
* **Phase 2 – End-to-end validation (multi-AoA chamber, e.g. feIFF)**
	+ **Simplified CDL-C channel matched to probe layout.**
	+ **Provides true multi-directional arrivals for natural Rx-beam sweeping.**
	+ **Confirms that the complete system, AI/ML model plus UE hardware, meets accuracy and responsiveness targets.**
 |
| [**R4-2513660**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_116bis/Docs/R4-2513660.zip) | CMCC | ***Proposal 1: for reported RSRP prediction, the related test apply to top-K beams, where K is 1 for absolute RSRP accuracy requirement verification, and K is larger than 1 for relative accuracy requirements verification.******Proposal 2: for successful rate for the correct prediction, for the FFS on top-N (N</=K) predicted beams have to be considered in the success rate evaluation, it is proposed that whether N value can be larger than 1 is related with x value.*** * ***Either x =0 and N>1, or x of non-zero value and N=1 are OK;***
* ***It is not necessary that N>1 and x of non-zero value are considered together***

***Proposal 3: for the case that both RSRP and beam ID are reported, it is proposed that both RSRP accuracy requirements and successful rate for the correct prediction are applied.*** ***Proposal 4: for relative RSRP accuracy, it is proposed that the reported L1-RSRP cannot be measured RSRP, since RAN1 agreed that the RSRP of predicted beam(s)in the report of inference results is the predicted RSRP which is based on AI/ML output.******Proposal 5: for relative RSRP accuracy, it is proposed that beam index n owns the largest reported value.*** ***Proposal 6: for BM case 1, it is proposed that*** * ***The relative RSRP accuracy for reported beams during inference reporting = (predicted L1-RSRP of beam index i - predicted L1-RSRP of beam index n) - (ground truth of L1-RSRP of beam index i - ground truth of L1-RSRP of beam index n), ~~[~~where the beam index n owns the largest reported value~~]~~***

***Proposal 7: for BM case 2, it is proposed that*** * ***The relative RSRP accuracy for reported beams during inference reporting = (predicted L1-RSRP of beam index i for time instance m - predicted L1-RSRP of beam index n) - (ground truth of L1-RSRP of beam index i for time instance m - ground truth of L1-RSRP of beam index n), where the beam index n owns the largest reported value among all the predicted beams. 1<=m<=M where M is the number of time instance***
 |
| [**R4-2513691**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_116bis/Docs/R4-2513691.zip) | Qualcomm Incorporated | **Observation 1: The key requirements for AI/ML BM test setup are listed following:*** **Requirement#1: Channel model in the test**
	+ **The test should reflect performance in the field. Hence, CDL or simplified CDL to properly verify beam management performance.**
* **Requirement#2: Avoid the cheating**
	+ **Progress A: Model training for Set A’s is generated with channel parameters #1. Set B in UE measurement should have enough randomness compared to Set A’.**
	+ **Progress B: Set A is generated to make sure to use the same parameters (initial phase, angular spread, delay spread) as the Set B in the Progress A measurement step. NW antenna/beam configurations for Set A and Set B should be deterministic. For some prediction metrics (e.g., option 2 and 3), UE can pass the test without actual measurement if TE does not randomize the transmission of its beams.**
* **Requirement#3: Support the beam number of Set A**
	+ **TE needs to support Set A for ground truth verification.**

**Observation 2: To generate Set A and Set B, multiple AoAs/probes test setup is necessary.****Observation 3: Using Multiple AoA test setup with FR2 RRM test can reuse the existing FR2 test setup to save the test cost and a simplified CDL channel model needs to be specified.****Observation 4: 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 5: The AoDs/ZoDs of very weak clusters are unlikely to influence the selection of the strongest beam in any reasonably designed codebook of Set A and Set B.****Observation 6: 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.****Propsoal 1: RAN4 to consider the following options to define the simplfied CDL channel model with eIFF test setup:*** **Revmoe the weak clusters that would not influence the UE beam management**
* **Merge the clusters that have the same/similer AoAs**
* **Reduce values of per-cluster parameters such as CASD, CASA, CZSD, CZSA**

**Propsoal 2: RAN4 to evlaute whether the new simiplied CDL channel model generated from eIFF system can be used to properly verify AI/ML BM performance through AIML BM simulations.** **Proposal 3: RAN4 to investigate whether measurement uncertainty should be considered when defining AI/ML BM core requirements.** |
| [**R4-2513786**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_116bis/Docs/R4-2513786.zip) | Samsung | *I-BM model inference delay requirement***Proposal 1:** The absolute RSRP accuracy requirement shall only be applied to Top-K of predicted beams: * K = 1, at least for Rel-19 progress.

**Proposal 2:** The relative RSRP accuracy requirement shall only be applied to Top-K of predicted beams: * K = 2, i.e., relative RSPR between Top-1 and Top-2 beam, at least for Rel-19 progress.

**Proposal 3:** For relative RSRP accuracy requirement shall be updated as below: * Relative RSRP accuracy for reported beams during inference reporting = (predicted L1-RSRP of beam index i - reported L1-RSRP of beam index n) - (ground truth of L1-RSRP of beam index i - ground truth of L1-RSRP of beam index n), where the beam index n owns the largest reported value given by absolute RSRP reporting.
	+ Reported L1-RSRP can be predicted,
	+ Reported L1-RSRP can also be measured RSRP if Set B is subset of Set A
* Relative RSRP accuracy requirements apply for Case 1, and also for Case 2

**Proposal 4:** For the beam ID prediction accuracy in Rel-19, only top-1 (i.e., Top-N, N=1) predicted beam is considered in the successful rate evaluation/requirement.**Proposal 5:** The beam ID prediction accuracy shall be updated as: The successful rate for the correct prediction, * + The correct prediction is considered as maximum ground-truth RSRP among top-K predicted beams larger than or equal to the ground-truth RSRP of the strongest genie-aided beam(s) – x dB,
	+ ~~FFS on top-N (N</=K) predicted beams have to be considered in the success rate evaluation~~
		- ~~FFS on N values~~
	+ FFS on x and K values in performance part

*AI-BM TCI state switching delay requirement***Proposal 6:** Confirm the below condition for determining the target TCI state is known shall applied for both AI-BM case 1 and case 2: - The TCI state switch command is received within 1280 ms upon the last transmission of the RS resource for beam reporting or measurement and the UE has sent at least one L-1 RSRP report of the target TCI state (clause 8.10.2 of 38.133)*LCM related core requirement***Proposal 7:** RAN4 shall revisit the agreement that “UE has to be ready to start measurements for inference after sending RRC reconfiguration complete”, by considering UE can report configuration inapplicable and doesn’t set the release flag.**Proposal 8:** RAN4 may not need to define AI functionality activation delay requirement in Rel-19, unless * RAN4 reach the agreement on the starting and ending points for activation delay procedure
* RAN4 has clear implication of relevant requirement on UE implementation to load practical AI/ML model

*Performance monitoring core requirement***Proposal 9:** RAN4 can define performance monitoring requirement in Rel-19 only if * RAN4 can agree on the testing condition in which expected performance monitoring result is known, i.e., a condition in which UE is expected to have perfect good inference condition and then the expected performance monitoring result is Np = N
 |
| [**R4-2513806**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_116bis/Docs/R4-2513806.zip) | OPPO | **Proposal 1: For BM testability, RAN4 needs to:**1. **Clarify the limitations regarding the FR2 beams or probes that TE vendors could support**
2. **To determine the test setup for BM, a potential approach could be:**
* **Assuming that TE supports X probes in a chamber**
* **By utilizing X probes, X different angles of arrival (AOA) can be achieved and can be mapped to a TX beam transmission with X paths(clusters)**
* **Different TX beams in BM set A or set B can be represented by setting different pathloss for the different AOA paths, to mimic the beam sweeping procedure**
* **FFS the number of probes that could be supported by TE vendors**
* **FFS how to map the X AoAs and corresponding X pathloss to different beams in BM set A/B**

**Proposal 2: Further consider how to ensure that the BM model constructed on the DUT side can match and be utilized in the testing environment on the TE side.****Observation 1: The over-simplified OTA assumptions being used for BM may offer little practical guidance for field deployment.****Proposal 3: For 6G, RAN4 could explore whether/how to decouple the signal acquisition testing (e.g., for model inputs) from model performance evaluation (e.g., for model performance) for AI use cases.** **Proposal 4: Signal acquisition could continue to be tested via OTA, e.g., evaluating BM Set B measurement accuracy.** **Proposal 5: Model performance could be assessed using standardized datasets defined by RAN4 (RAN4 specified datasets).** |
| [**R4-2514057**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_116bis/Docs/R4-2514057.zip) | Ericsson | **Observation 1: Measurement delay may depend on various configurations and applied channel. RAN4 shall define a single number/value with respect to a common configuration.****Observation 2: RAN4 to confirm that inference time is the delay between the most recent measurement RS occasion and the time instant when the inference becomes valid.****Observation 3: Prediction only applies for MAC-CE based and RRC based TCI state switching, with prediction based known/unknown condition, DCI based TCI state switching delay isn’t impacted.****Observation 4: The TCI state switch delay may account for the predicted Tx beam in Set A during the switching period. Its transmission can be summarized as follows:*** **The predicted Tx beam in Set A shall be transmitted/configured for UE to measure, if the predicted Tx beam in Set A is the first SSB and/or the RS for L1-RSRP, otherwise**
* **the predicted Tx beam in Set A doesn’t need to be transmitted/configured for UE to measure, if the predicted Tx beam in Set A isn’t the first SSB and/or the RS for L1-RSRP.**

**Observation 5: For the case that only beam ID is reported, the order of top-K beams indicates the order of RSRP or the probability of each beam in Set A to be the Top-1 beam of implicitly.** **Observation 6: Only considering the baseline:*** **Top-K/1 (%), and**
* **The correct prediction is considered as maximum ground-truth RSRP among top-K predicted beams larger than or equal to the ground-truth RSRP of the strongest genie-aided beam(s) – x dB**

**may result in a wrong, e.g., bad quality, beam being adopted by NW, and results in degraded performance than the legacy measurement solution.****Observation 7: From the network perspective, obtaining a plurality of predicted beams, each subject to specific prediction accuracy requirements,** **rather than only one or some of them, is advantageous for enhancing reliability and robustness in beam management.****Observation 8: For AI/ML model supporting RSRP and beam ID reporting, if only RSRP accuracy is required, an underlying problem is that the Top-1/top-K predicted beam is far away from the Top-1/top-K genie-aided beam since no comparison to genie-aided beam** **is performed.****Observation 9: For the** **case of the AI/ML model with label (beam ID) + RSRP output type, fulfilling beam prediction accuracy and RSRP accuracy may result in more stringent requirements compared with the case of the AI/ML model only with label (beam ID) output type.** **Proposal 1: Measurement delay is defined as follows:*** **For BM-case1, Measurement delay is one shot L1-RSRP measurement delay.**
* **For BM-case2, Measurement delay is 4 \* (L1-RSRP measurement delay).**

**Proposal 2: RAN4 shall define a reference configuration, e.g., reference signal configurations, UE movement, channel for test and so on, to verify the prediction accuracy based on the agreed measurement delays.****Proposal 3: The parameter d' for UE CSI computation in TS38.214 may be regarded as the inference time.** **Proposal 4: For aperiodic report, RAN4 to define the overall prediction reporting delay, which shall be defined in TS38.133, at least includes measurement delay + reporting delay. Wherein,** * **Measurement delay: shall be specified by RAN4, and**
* **Reporting delay: as defined in Clause 9.5.3.3 ‘Aperiodic Reporting’ in TS38.133 (without modification), which doesn’t need to restated.**

**Proposal 5: For semi-persistent and periodic report, RAN4 to define the overall prediction reporting delay, which shall be defined in TS38.133, includes measurement delay +inference delay+time for the first available reporting occasion. Wherein,** * **Measurement delay: shall be specified by RAN4, and**
* **Inference time: shall be specified by RAN4, as per UE capability, and**
* **Time for the first available reporting occasion: as defined in Clause9.5.3.1 ‘Periodic Reporting’ and Clause9.5.3.2‘Semi-Persistent Reporting’ in existing TS38.133, which doesn’t need to restated.**

**Proposal 6: RAN4 to clarify that the use case of prediction in TCI state is:*** **The RS resource for prediction is the RS in target TCI state or QCLed to the target TCI state.**

**Proposal 7: The term “detectable” in the condition ‘The SSB associated with the TCI state remain detectable’ should be clarified as follows:*** **If the RS resource for prediction is the SSB associated with the TCI state, the (predicted) SSB associated with the TCI state remain detectable based on prediction.**
* **If the RS resource for prediction isn’t the SSB associated with the TCI state, the (measured) SSB associated with the TCI state remain detectable based on measurement.**

**Proposal 8: Update known condition to one of the following options:****Option 1: ‘TCI state switch command is received within 1280 ms upon the last transmission of the RS resource for beam predict reporting or measurement for prediction.’****Option 2: ‘TCI state switch command is received within 1280 ms upon the last transmission of the RS resource for beam predict reporting ~~or measurement for prediction.’~~****Proposal 9: RAN4 to confirm that prediction only impacts the definition of known conditions in Clause 8.10.2 of TS38.133, and does not impact the fundamental delay requirements specified for MAC-CE based TCI state switching in Clause 8.10.3 of TS38.133 and for RRC based TCI state switching in Clause 8.10.4 of TS38.133.****Proposal 10: RAN4 to confirm that prediction based known conditions for TCI state are also applicable for Clause 8.10.3A ‘MAC-CE based TCI state switch delay in HST FR2 scenarios’ of TS38.133.****Proposal 11: Enhance the baseline metric for beam prediction by testing both cases: K = 1 and K = 2 or 3, with the beam prediction metrics baseline as follow:*** **‘Top-K/1 (%)’, and**
* **‘The correct prediction is considered as maximum ground-truth RSRP among top-K predicted beams larger than or equal to the ground-truth RSRP of the strongest genie-aided beam(s) – x dB’.**

**Proposal 12: The parameter x in beam prediction accuracy metrics should be set to a value that reflects the expected magnitude of measurement error and other errors, [1] dB may be taken as a baseline.****Proposal 13: 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 is adopted.**
* **In case of the AI/ML model with label (beam ID) + RSRP output type**
	+ **Beam prediction accuracy and RSRP accuracy, are adopted simultaneously.**

**Proposal 14: For the absolute RSRP, the index i shall be any beam index in all beams based on predicted L1-RSRP as per UE reporting, i.e.,*** **The absolute RSRP accuracy = predicted L1-RSRP of beam index i – groundtruth of L1-RSRP of beam index i. The index i may be any beam index in top-K beams based on prediction report.**

**Proposal 15: Relative RSRP accuracy for reported beams during inference reporting = (predicted L1-RSRP of beam index i - predicted L1-RSRP of beam index n) - (ground truth of L1-RSRP of beam index i - ground truth of L1-RSRP of beam index n), [where the beam index n owns the largest reported value].****Proposal 16: RAN4 to define prediction monitoring requirements for performance monitoring Type 1 Option 2 (UE-assisted performance monitoring)****Proposal 17: RAN4 shall take the below RAN1 agreements into performance monitoring requirements:*** **Evaluation period for performance monitoring:**
	+ **N transmission occasion of the CSI-RS/SSB resources for monitoring, at least for BM-case 1, Where N = 1, 3, 7, 15 is configured in CSI-ReportConfig.**
* **Report quantity RS-PAI,**
	+ $ N\_{p}$ **is defined as the total count of accurate reference signal prediction instance(s) that meets the condition, among N latest transmission occasion(s).**

**Proposal 18: RAN4 to study the two questions regarding multiple/single AoA:*** **What is the (maximal) difference between really tested Tx/Rx beam and the artificial Tx/Rx beam?**
* **If the difference can impact AI/ML model and the prediction accuracy?**

**Proposal 19: RAN4 to check if the 32 beams after simplified CDL channel can be produced by [8] (if it is feasible to be realized in test bed) probes which may achieve the balance between performance and implementation complexity.****Proposal 20: 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.**
 |
| [**R4-2514092**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_116bis/Docs/R4-2514092.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: In current RAN4 spec, we do not observe the limitation on absolute RSRP accuracy.****Proposal 3: RAN4 shall consider absolute RSRP accuracy requirement applying to Top-1 of predicted beams and other beams.****Observation 6: The whole principle for defining relative RSRP accuracy is that the predicted L1-RSRP minus ground truth beams which is similar to legacy definition.** **Observation 7: If the reported L1-RSRP is reported measured RSRP, this components will disappear based on current formulation.****Proposal 4: The reported L1-RSRP shall be reported predicted L1-RSRP not the measured RSRP.****Observation 8: Top-K/1 is the relaxation of Top-1/1, UE vendors has extended margin on this performance metric and the prediction accuracy can be up to 97% even the measurement error has been added.****Observation 9: If the margin is set too large, the threshold for performance validation becomes low enough that any UE can pass the test, leading to the realization that AI’s predictive capabilities are actually not as good as legacy methods.****Proposal 5: At least x=0 shall be considered in performance metrics.****Proposal 6: The concrete value of x when x>0 shall be defined based on simulation results.****Proposal 7: The value of x shall not be set too large in order to guarantee the AI performance test and prevent UE easily passes the test.** **Observation 10: the maximum ground truth beam among Top-K beams has small difference or equal to Top-1 beam if only Top-1 beam is considered. The test will definitely pass. Nonetheless, if the rest of beams in Top-K predicted beams are worse beams, the whole AI/ML performance will be wrongly judged as successful.****Proposal 8: RAN4 shall support to introduce Top-N beams for beam ID prediction to ensure high reliability of AI/ML.****Proposal 9: Different interpretation of Top-K/N% could be seen as below:**1. **One of Top-N beams contains in Top-K predicted beams; (like RAN1’s agreement on performance monitoring)**
2. **All of Top-N beams contains in Top-K predicted beams with corresponding order.**
3. **All of Top-N beams contains in Top-K predicted beams without corresponding order.**

**Observation 11: For interpretation A, the test will be passed easily if only one of Top-N in Top-K predicted beams.****Observation 12: For interpretation B, compared to interpretation A, this metric will not cause test passes easily, but this metric has large probability that the test will not be passed since the conditions are quite strict.****Observation 13: For interpretation C, only guarantee all Top-N beams contain in Top-K beams, Top-N beams are measured by UE which can be assumed as groundtruth beams. That is, if all groundtruth beams contain in AI/ML predicted Top-K beams, the test will be passed.****Proposal 10: The definition of Top-K/N% shall be all of Top-N beams contains in Top-K predicted beams without corresponding order.****Proposal 11: 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 12: For measurement delay, the legacy L1-RSRP measurement period shall be reused.****Observation 14: The total measurement samples may refer to the number of predicted future time instances. If T=1, the measurement samples may be too less to predict more accurate prediction of future time instance. If T=8, the measurement samples may be too much and the training model will be over-fitting.****Proposal 12: For measurement delay in case 2, the T value could follow the number of future time instances proposed by RAN1, which is [2 or 4].****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 15: The smaller SNR conditions, the larger measurement error.****Observation 16: The performance degradation will be occurred due to added measurement error.****Observation 17: For Top-1, the prediction accuracy is below 86% even if there is no error on test data.****Observation 18: 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.** |
| [**R4-2514248**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_116bis/Docs/R4-2514248.zip) | Rohde & Schwarz | **Proposal 1:** **(feIFF Setup)** The multi-AoA IFF based test setup with 4x active cross-polarized probes (CATRs) placed in a single plane and separated by 30º, 60º and 60º respectively, is defined as feIFF (further enhanced IFF) test setup. The concrete probe layout and the alignment with the coordinate system to be used for testing is FFS.A screen shot of a video game  Description automatically generatedA screen shot of a cell phone  Description automatically generated**Observation 1:** **(feIFF Setup)** The feIFF test setup supports 4 active cross-polarized probes with large distinctive angular separation of up to 150°, overall big quiet zone of at least 30 cm, lowest MUs and ensures preservation of (RF-RRM) side conditions. **Observation 2:** **(feIFF Setup)** The feIFF test setup is a feasible upgrade (superset) of the existing legacy eIFF setup. As such, if feIFF is concluded as capable for AI/ML BM testing, a single system covers the whole conformance testing, ensuring continuity in the cost sensitive FR2 testing ecosystem. **Observation 3:** **(Simplified Channel)** CDL channel model can be simplified by merging the clusters that have the same AoA and removing the weak clusters falling outside the UE’s spherical coverage (i.e. < -10dB) resulting in the simplified channel model in Table 1.**Observation 4:** **(Simplified Channel)** The required total AoA spread of 150° of the simplified channel in Table 1 can be emulated with the feIFF probe layout. Table 1 – Channel model parameters for UMi CDL-C at 28 GHz after merging clusters with similar AoAs and removing weak clusters

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Cluster # | Absolute Delay [ns] | Power in [dB] | AOD in [°] | AOA in [°] | ZOD in [°] | ZOA in [°] |
| 1 | 0 | -7.4318 | -30.4353 | -134.4434 | 98.9242 | 83.3318 |
| 2 | 12.594 | -1.2500 | -20.9269 | 129.1633 | 99.1915 | 72.5229 |
| 5 | 13.056 | -5.5318 | -28.0782 | -152.8206 | 99.5732 | 71.1282 |
| 6 | 38.196 | 0.0000 | -11.6982 | 164.1145 | 99.306 | 74.7544 |
| 13 | 73.71 | -8.1318 | -33.911 | 93.1719 | 100.165 | 85.4238 |
| 14 | 78.498 | -9.8318 | -37.5066 | -112.0441 | 100.2604 | 64.1548 |
| Per-Cluster Parameters |
| Parameter | CASD in [°] | CASA in [°] | CZSD in [°] | CZSA in [°] | XPR in [dB] |  |
| Value | 0.799 | 10.4021 | 0.5726 | 4.8814 | 7 |  |

**Observation 5: (Simplified Channel)** The simplified channel model parameters for UMi CDL-C can be implemented in the candidate test system feIFF, with 3 possible options (Figure 3): **Option A:** with flattened elevation (Table 2) and strongest cluster aligned to one probe; **Option B:** with flattened elevation (Table 2) and mid-point of the total angular spread of the clusters is aligned to the mid-point of the spread of the probes; **Option C:** with flattened elevation, AoAs aligned to probe layout and no intra‑cluster angle spread (Table 3).Table 2 – Channel model parameters for UMi CDL-C at 28 GHz with flat ZoA

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Cluster # | Absolute Delay [ns] | Power in [dB] | AOD in [°] | AOA in [°] | ZOD in [°] | ZOA in [°] |
| 1 | 0 | -7.4318 | -30.4353 | -134.4434 | 98.9242 | 74.51134 |
| 2 | 12.594 | -1.2500 | -20.9269 | 129.1633 | 99.1915 | 74.51134 |
| 5 | 13.056 | -5.5318 | -28.0782 | -152.8206 | 99.5732 | 74.51134 |
| 6 | 38.196 | 0.0000 | -11.6982 | 164.1145 | 99.306 | 74.51134 |
| 13 | 73.71 | -8.1318 | -33.911 | 93.1719 | 100.165 | 74.51134 |
| 14 | 78.498 | -9.8318 | -37.5066 | -112.0441 | 100.2604 | 74.51134 |
| Per-Cluster Parameters |
| Parameter | CASD in [°] | CASA in [°] | CZSD in [°] | CZSA in [°] | XPR in [dB] |  |
| Value | 0.799 | 10.4021 | 0.5726 | 0 | 7 |  |

Table 3 – Channel model parameters for UMi CDL-C at 28 GHz with AoA aligned to probe layout and no intra-cluster angle spread

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Cluster # | Absolute Delay [ns] | Power in [dB] | AOD in [°] | AOA in [°] | ZOD in [°] | ZOA in [°] |
| 1 | 0 | -7.4318 | -30.4353 | -114.436 | 98.9242 | 74.51134 |
| 2 | 12.594 | -1.2500 | -20.9269 | 125.5639 | 99.1915 | 74.51134 |
| 5 | 13.056 | -5.5318 | -28.0782 | -174.436 | 99.5732 | 74.51134 |
| 6 | 38.196 | 0.0000 | -11.6982 | -174.436 | 99.306 | 74.51134 |
| 13 | 73.71 | -8.1318 | -33.911 | 95.5639 | 100.165 | 74.51134 |
| 14 | 78.498 | -9.8318 | -37.5066 | -114.436 | 100.2604 | 74.51134 |
| Per-Cluster Parameters |
| Parameter | CASD in [°] | CASA in [°] | CZSD in [°] | CZSA in [°] | XPR in [dB] |  |
| Value | 0.799 | 0 | 0.5726 | 0 | 7 |  |

A circular graph with different colored circles  Description automatically generatedFigure 3 – Simplified UMi CDL-C at 28 GHz alignment to the candidate test system feIFF**Observation 6: (Simplified Channel)** The simplified channel model parameters for UMi CDL-C do not affect the doppler per cluster, i.e. the latter is still emulated by the test system. **Proposal 2: (Simplified Channel)** Consider the channel models described in Table 2 (Option A/B) and Table 3 (Option C) as candidates for CDL-based simplified channel models for multi-AoA testing of AI/ML BM.**Proposal 3: (Channel Evaluation)** Evaluate the suitability of the simplified channel for AI/ML BM testing through system level simulations. Compare the prediction performance between simulations with the full channel and the simplified channel, when using the same training model and performance metric.**Observation 7: (Channel Evaluation)** Based on preliminary simulations with 8 Set B & 32 Set A beams, there is only a negligible effect (deviation of max ~1%) on the prediction accuracy for Top 3 & Top 1 beams between the cases with full CDL channel model (Reference - 24 clusters) and simplified CDL channel model suitable for feIFF (Options A/B/C - 6 clusters).**Observation 8: (Channel Evaluation)** Based on preliminary simulation results, the simplified channel models for feIFF in Table 2 and Table 3 can be considered for further evaluation for AI/ML BM testing. **Proposal 4: (Channel Evaluation)** RAN4 to start the evaluation of CDL-C simplified channel models for feIFF with 6 Clusters in Table 2 and Table 3 through AI/ML BM system level simulations. Evaluate the channel suitability as per the approach described in Proposal 3. |
| [**R4-2514308**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_116bis/Docs/R4-2514308.zip) | Nokia | **Observation 1:** Performance evaluation of beam prediction functionality considers all the reported top-K beams.**Observation 2:** Prediction of $i^{th}$ ranked predicted beam is deemed a successful event, if RSRP of $i^{th}$ ranked predicted beam, $\hat{y}\_{i}$ is greater than or equal to the RSRP of the corresponding $i^{th}$ ranked genie-aided beam, $y\_{i}-x$**Proposal 2: RAN4 to discuss and agree on the FFS components of the RAN4#114bis agreement on the beam Id prediction metric considering the following:**1. At least $N$ strongest ground-truth RSRP beams among the top-$K$ predicted beams must be included in the success rate evaluation, where $2\leq N\leq K$
	* RAN4 should consider $N=2$, if there is no consensus in RAN4 for $N>2$
2. Correct prediction of these $N$ beams is evaluated using a fixed value of $x$

**Proposal 3: RAN4 to discuss and incorporate the following changes to complete the description of the beam Id prediction metric:**1. The correct prediction of ith ranked predicted beam, $s\_{i}$, is considered as the ground truth RSRP of the ith ranked predicted beam larger than or equal to the ground truth RSRP of the ith ranked genie-aided beam – x dB, where $i=1, 2,…,N$. The ranking of beams within both the top-K predicted and genie-aided beam sets is based on ground-truth RSRPs

$$s\_{i}=l\left(\left|\hat{y}\_{i}-y\_{i}\right|\leq x\right)$$1. **Compute the success rate,**$ s$**, by aggregating the correct prediction of the strongest N beams among the top-K predicted beam set**

$$s=\frac{1}{N}\sum\_{i=1}^{N}s\_{i}$$**Observation 3:** There is no need to have separate requirements for prediction reporting delay and measurement period for inference for AI/ML BM use cases.**Proposal 4: Regarding the value of T in the delay requirement for predicted L1-RSRP, RAN4 should consider T=1 for all types of reports i.e. periodic, semi-persistent and aperiodic.****Proposal 5: In case if the inference delay is negligible, RAN4 should consider legacy delay requirements for the predicted L1-RSRP as well.****Proposal 6: In case if the inference delay is not negligible for the definition of the delay requirement for predicted L1-RSRP, RAN4 needs to further discuss and agree on the value of the inference delay.****Proposal 7: RAN4 needs to define accuracy requirements for UE-assisted monitoring where the UE reports the beam prediction accuracy, RS-PAI, in BM-Case1 and BM-Case2. The accuracy of RS-PAI should be tested at least in static radio conditions.****Proposal 8: The suggested changes in the draft CR (R4-2514311) should be endorsed to update known conditions for TCI state in clause 8.10.2 of TS 38.133.** |
| [**R4-2514311**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_116bis/Docs/R4-2514311.zip) | Nokia | Draft CR |
| [**R4-2514457**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_116bis/Docs/R4-2514457.zip) | Qualcomm Incorporated | **Observation 1:** The FFS of N > 1 in beam ID accuracy prediction metric can be expanded to different metrics.* Potential metric 1: The metric could check if N out of top K predicted beams are contained among the M strongest genie aided beams.
* Potential metric 2: The metric could also check if any N out of top K predicted beams are within X dB of any two of the M strongest genie aided beams.
* Potential metric 3: The metric could also check if the ground truth RSRP of top N predicted beams are within X dB of the ground truth RSRP of the two genie aided strongest beams.

**Observation 2:** The proponents of N > 1 claim that the current version of the agreement only checks if one of top K predicted beams is within X dB margin, in terms of ground truth RSRP, of the genie aided strongest beam. Network does not know which of the top K predicted beams is within X dB margin of the genie aided strongest beam. **Observation 3:** The proponents N > 1 also claim that it gives network idea about UE’s capability to predict and report multiple strong beam IDs.**Observation 4:** If K > 1, network must transmit the top-K predicted beams in a 2nd round of beam sweep and collect UE’s report regarding the strongest beam among top-K predicted beams before changing active TCI state.**Observation 5:** Currently agreed procedures do not allow network to dynamically configure (e.g., via DCI or MAC-CE) the 2nd round of beam sweep based on UE’s inference results regarding top-K predicted beams.**Observation 6:** The issue of 2nd round of beam sweep, mentioned in observation 4 and 5, can be avoided by making K = 1.**Observation 7:** In FR2, networks typically cannot communicate with multiple UEs via the same beam. The knowledge of the 2nd strongest, defined via ground truth RSRP, predicted beam may not be essential for the network.**Observation 8:** Beam index n, owning the largest reported value, can be an element of setB. However, UE reports it as predicted L1-RSRP while reporting it as part of a top beam of setA. It does not matter whether the reported and predicted L1-RSRP of this beam is exactly equal to the measured L1-RSRP of this beam.**Observation 9**: Channel may vary significantly across different prediction time instances. It is not meaningful to compare relative accuracy of reported L1-RSRP across two different prediction time instances.**Observation 10:** RAN4 has kept the required minimum number of observation instances as FFS in the evaluation period definition of case 2.**Observation 11:** RAN1 has recently agreed to allow UE, at least, K consecutive observation occasions before UE is required to transmit a CSI report for inference, where K would be indicated by a new UE capability.**Observation 12:** RAN4 has agreed to the following regarding the QCL relationship of TCI state handling in AI-ML BM.* If the predicted TX beam in setA is QCL Type-D to a known measured Tx beam, the corresponding Rx beam is known.
* If the predicted Tx beam in set A is not QCL Type-D to a known Tx beam, known TCI state conditions will be based on UE capability.

**Observation 13:** In legacy RRM spec, a TX beam is classified as “known” if the UE has measured it during the last 1.28 seconds and if UE has sent a report regarding that beam.**Observation 14:** UE measures setB beams to generate and send inference report for setA but UE may not send measurement reports of setB to the network.**Observation 15:** RAN1 has agreed to increase the minimum time separation between reference signal and aperiodic L1-RSRP report by d’ in AI-ML beam prediction.* Exact value of d’ is TBD in RAN1.

**Observation 16:** RAN4 has agreed to include inference delay in prediction delay of AI-ML BM. This is equivalent to the parameter d’ of RAN1 spec.**Proposal 1:** Update the previous agreement with the following changes: * “The successful rate for the correct prediction,
	+ The correct prediction is considered as ~~maximum~~ ground-truth RSRP ~~among~~ of top-~~K~~1 predicted beam~~s~~ being larger than or equal to the ground-truth RSRP of the strongest genie-aided beam(s) – x dB,
	+ ~~FFS on top-N (N</=K) predicted beams have to be considered in the success rate evaluation~~
		- ~~FFS on N values~~
	+ FFS on x values
* Note: if x=0 ~~and N=1~~, it means Top-~~K~~1/1 (%) : the percentage of "the Top-1 strongest beam is ~~one of~~ the Top-~~K~~1 predicted beam~~s~~"

**Proposal 2:** RAN4 uses the agreed baseline for relative L1-RSRP of case 1 with the following changes:* + Relative RSRP accuracy for reported beams during inference reporting = (predicted L1-RSRP of beam index i – ~~reported~~ predicted L1-RSRP of beam index n) - (ground truth of L1-RSRP of beam index i - ground truth of L1-RSRP of beam index n), ~~[~~where the beam index n owns the largest reported value~~]~~.
		- ~~Reported L1-RSRP can be predicted,~~
		- ~~FFS whether reported L1-RSRP can be measured RSRP~~

**Proposal 3:** Apply proposal 3 to case 2 where predicted and ground truth L1-RSRP of both beam index i and n correspond to the same time instance. In other words, define relative L1-RSRP accuracy in the following way for case 2: * + Relative RSRP accuracy for reported beams during inference reporting = (predicted L1-RSRP of beam index i at time instance t – predicted L1-RSRP of beam index n at time instance t) - (ground truth of L1-RSRP of beam index i at time instance t- ground truth of L1-RSRP of beam index n at time instance t), where the beam index n owns the largest reported value at time instance t.

**Proposal 4:** T in the measurement period for inference in case 2 of RAN4 agreement is equal to the RAN1 defined UE capability K, where K denotes the number of consecutive transmission occasion that UE needs to receive before transmitting a CSI report for inference in BM-Case 2.**Proposal 5:** RAN4 clarifies that “known measured TX beams” include setB beams in the agreement on QCL relationship on TCI state handling.**Proposal 6:** The inference timelines mentioned in RAN4 spec should be aligned with d’ in RAN1 spec.* Note 1: d’ is the extension in time separation between RS and AP L1-RSRP report in AI-ML based beam prediction.
* Note 2: d’ can vary between spatial only and spatio-temporal beam prediction. Exact values of d’ will be discussed on August 2025 meeting in RAN1.
 |
| [**R4-2513219**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_116bis/Docs/R4-2513219.zip) | CATT | **Observation 1: Different performance requirements need to be separately defined for beam ID prediction and RSRP prediction.** **Proposal 1: For AI/ML based beam management with both predicted RSRP and beam ID reported, two options can be considered for defining performance requirements:** * **Option 1: The legacy L1-RSRP performance requirements are reused, regardless of overhead reduction ratio.**
* **Option 2: The legacy L1-RSRP performance requirements are relaxed with margins, and larger margin values are used for higher overhead reduction ratios.**

**Proposal 2: For AI/ML based beam management with predicted beam ID reported only, RAN4 can define a lower bound for the successful rate as performance requirement regardless of the values of K, N, x and SINR.** |
| [**R4-2513415**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_116bis/Docs/R4-2513415.zip) | vivo | **Proposal 1: For RRM performance requirements for beam management, RAN4 to consider the following requirements；*** **Beam prediction accuracy**
* **Absolute Predicted RSRP accuracy**
* **Relative Predicted RSRP accuracy**
* **Accuracy requirements for performance monitoring(to verify whether the UE correctly reports the number of accurate predictions out of N inference attempts.)**

**Proposal 2: For beam prediction accuracy and RSRP accuracy, when defining specific performance requirements, it is necessary to determine how many different sets of metric requirements should be defined to accommodate various scenarios. RAN4 to first identify the factors that influence prediction performance, potentially including:*** **The number of set B beams, the number of set A beams**
	+ **candidate value:**
		- **With measurements of Set B of beams that of 1/4 of Set A of beams (e.g., 8 for set B, 32 for set A)**
		- **With measurements of Set B of beams that of 1/8 of Set A of beams (e.g., 8 for set B, 64 for set A)**
* **Beam pattern**
	+ **Consider the beam pattern defined in Table 6 of the existing simulation assumptions [1]**
* **Prediction scenario:**
	+ **Wide beam to narrow beam (SSB to CSI-RS)**
	+ **narrow to narrow (CSI-RS to CSI-RS)**
* **The value of N**
	+ **candidate value: 1,2,3,4**
* **The value of X**
	+ **0,1,2,3**
	+ **Specific values should be determined based on N and simulation results**

**Proposal 3: For test cases for AI/ML based beam management, RAN4 at least to define RRM test cases for the following requirements:*** **Prediction delay requirement**
 |
| [**R4-2513423**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_116bis/Docs/R4-2513423.zip) | Huawei,HiSilicon | **Observation 1:** With Gaussian distribution assumption of baseband measurement error, under different SNRs, the mean values are all close to 0, while the variance increases as the SNR drops.**Proposal 1:**The simulation results for different prediction accuracy regarding to KPI 1 is shown in Table 1.**Table 1. Measurement error impact on prediction accuracy regarding to KPI 1**

|  |  |  |
| --- | --- | --- |
| KPI-1 | Model input w/o measurement error | Model input w measurement error |
| maximum RSRP among top-K predicted beams is larger than the RSRP of the strongest beam – x dB | x = 1dB | 83.02% | 60.75% |
| x = 2dB | 92.49% | 72.87% |
| x = 3dB | 96.29% | 81.84% |

**Observation 2:**The more stringent the KPIs are, the greater the impact of the measurement error on the prediction accuracy will be.**Proposal 2:** The simulation results for different prediction accuracy regarding to KPI 2 is shown in Table 2.**Table 2. Measurement error impact on prediction accuracy regarding to KPI 2**

|  |  |  |
| --- | --- | --- |
| KPI-2 | Model input w/o measurement error(case 1) | Model input w measurement error(case 2b) |
| 90%-tile L1-RSRP difference between the maximum RSRP of the Top-1/ Top-3/ Top-5 predicted beam(s) and the ground truth L1- RSRP of the genie aided strongest beam | Avg.  | CDF\_strong\_1 | 2.547dB | 5.871dB |
| CDF\_strong\_3 | 2.345dB | 5.392dB |
| CDF\_strong\_5 | 2.209dB | 4.893dB |
| w/o avg. | CDF\_strong\_Top1 | 2.547dB | 5.871dB |
| CDF\_strong\_Top2 | 2.380dB | 5.492dB |
| CDF\_strong\_Top3 | 2.139dB | 4.781dB |
| CDF\_strong\_Top4 | 2.049dB | 4.445dB |
| CDF\_strong\_Top5 | 1.967dB | 4.044dB |
| 90%-tile L1-RSRP difference between the predicted L1-RSRP of the Top-1/ Top-3/ Top-5 predicted beam(s) and the ground truth L1-RSRP of the same beams | Avg. | CDF\_same\_Top1 | 2.151dB | 7.250dB |
| CDF\_same\_Top1&2&3 | 2.236dB | 7.406dB |
| CDF\_same\_Top1&2&3&4&5 | 2.419dB | 7.444dB |
| w/o avg. | CDF\_same\_Top1 | 2.151dB | 7.250dB |
| CDF\_same\_Top2 | 2.200dB | 7.314dB |
| CDF\_same\_Top3 | 2.435dB | 7.702dB |
| CDF\_same\_Top4 | 2.674dB | 7.730dB |
| CDF\_same\_Top5 | 2.689dB | 7.287dB |

**Observation 3:** Regarding RSRP prediction accuracy, measurement errors have a larger impact on the Top 1 predicted beam than on the Top 5th predicted beam.**Proposal 3:** RAN4 define requirements with measurement error as model input. **Proposal 4:** RAN4 will not select KPI under which acceptable performance cannot be achieved, including* the KPI under which the prediction accuracy is less than 80%, e.g, Top K/1 with K<4 without RSRP margin.
* the KPI under which the 90%-tile L1-RSRP difference is larger than 5dB, e.g, maximum RSRP among top-K predicted beams is larger than the RSRP of the strongest beam – x dB with x<3dB.
 |
| [**R4-2513454**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_116bis/Docs/R4-2513454.zip) | Xiaomi | **Observation 1: The similar metric Top-K/M is agreed in RAN1 performance monitoring, where at least one of the Top M beam(s) of the resource set(s) for monitoring is among Top-K predicted beam(s).****Proposal 1: In order to align with RAN1, Beam index prediction definition should at least include original option 2:*** **For N value:**
	+ **When X is larger than 0 dB, N = 1.**
	+ **When X = 0dB, N = 1 or 2 depending on K value.**

**Observation 2: In legacy, absolute RSRP is only defined for the largest RSRP and relative accuracy is defined for other differential RSRP in the report.****Proposal 2: Only define Absolute RSRP accuracy requirement for Top-1 of predicted beam.****Proposal 3: For relative predicted RSRP accuracy, define unified definition for both BM-case 1 and BM-case 2:**$$Relative RSRP Accuracy\_{i,t}=\left(Predicted RSRP\_{i,t}-Predicted RSRP\_{max}\right)-\left(True RSRP\_{i,t}-True RSRP\_{max}\right)$$**Where** $Predicted RSRP\_{max}$ **is the largest RSRP across all time instance and beams in the report。**$Predicted RSRP\_{i,t}$ **is RSRP of beam index i at time t。**$True RSRP\_{max}$ **is the true RSRP corresponding to the beam/time of** $Predicted RSRP\_{max}$**。Where** $Predicted RSRP\_{i,t}-Predicted RSRP\_{max}$ **is predicted differential RSRP in the report.****Observation 3: Significant Performance Degradation in Low SNR Groups. The performance degradation is more significant in UE groups with lower maximum SNR.****Observation 4: Higher Stability in High SNR Groups. UE groups with higher maximum SNR (e.g., UE group 1 and UE group 2) show better performance stability under different measurement errors.****Observation 5: The SINR distribution for all UE groups spans a large range, and this distribution can significantly impact the final accuracy performance.****Proposal 4: RAN4 to align simulation results without measurement error first.****Proposal 5: To ensure consistent and reliable simulation results, we recommend that RAN4 align the following factors:*** **1. Method of Adding Measurement Error for Different TX Beams：**
	+ - **Generate error distributions separately for each individual Tx-Rx beam pair based on different SNRs and apply different Gaussian standard deviations to the corresponding Tx-Rx beam pair measurements in the SLS.**
* **2. Gaussian Standard Deviation for Different SNRs**
	+ - **Establish reference standard deviation values for critical SNR operating points.**
* **3. Largest SNR Distribution Among All Selected UEs**
	+ - **Divide the largest SNR range into several groups and define reference percentages for each group.**
* **4. SNR Distribution for TX Beams in Set B**
	+ - **Align the SNR distribution for all TX beams in set B to ensure consistency across different companies.**
 |
| [**R4-2513586**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_116bis/Docs/R4-2513586.zip) | Apple | 1. **If predicted RSRP values from AI/ML were guaranteed to have the same accuracy as legacy measured RSRP for every reported beam, then a separate beam-prediction accuracy metric would not be needed, because prediction and measurement would be equally reliable**
2. **Case3 Simulation with Noise-Aware Data Quality Control**RAN4 is invited to define beam-management (BM) performance requirements based on Case 3 of the agreed simulation framework:
* **Noise conditions**
	+ **Measurement error is present in training data, inference inputs, and ground truth used for training and testing.**
* **Noise-aware data quality control**
	+ **Model inputs: span a realistic SNR range to reflect deployment conditions.**
	+ **Target labels: collected at the highest practical SNR or by averaging repeated measurements to reduce label noise and allow for noise aware training**
* **Objective**
	+ **Ensure BM models are trained and tested under realistic noisy conditions while preserving clean, reliable ground truth for accurate performance evaluation.**
1. Meeting the FR2 Accuracy Target P90(|error|) ≤ 6 dB

**RAN4 is invited to adopt the following measures to reduce AI/ML beam-prediction error and meet the FR2 requirement:*** **Reduce SNR-sensitive errors by ensuring higher-SNR side conditions for the reference signals used in AI/ML inputs.**
* **Improve SNR through RS power boosting or 1–2 non-coherent repetitions.**
* **Revisit RF assumptions by defining deployment RF profiles with tighter variance where feasible.**
* **Apply noise-aware training with SNR-boosted labels to reject noise from inputs and improve model robustness, lowering the baseline prediction error.**
1. Beam-ID Prediction Accuracy Requirements per Reporting Case

**RAN4 should define accuracy requirements tailored to the selected reporting method:*** **Case 1 – Model-Ranked (inclusion and ranking order)**
	+ **Both beam inclusion and ranking matter.**
	+ **AI/ML uses its ranked beams to compare pairwise with the genie RSRPs within the allowed x dB, following the AI/ML-predicted beam order**
* **Case 2 – Probe-Ranked (inclusion only)**
	+ **Only beam inclusion and Top-K set correctness are required.**
	+ **Re-ordering of the Top-K beams takes place based on actual RSRP measurements, not on the AI/ML model’s predicted order.**
* **Action**
	+ **Define separate accuracy requirements for the two cases, or select one reporting case and set requirements accordingly.**
1. RAN4 to adopt a beam prediction accuracy metric based on a RSRP threshold model x = x₀ + Δ, where x₀ is the acceptable performance gap between predicted and genie beams, and Δ accounts for RF-induced uncertainty in RSRP measurements and it should be derived statistically from simulations modeling realistic RF error distributions and correlation across Tx beams.
2. We propose setting **N > 1** when evaluating Top-K beam prediction accuracy to reflect realistic gNB usage of multiple fallback beams. Considering more than one strong beam improves robustness, and avoids overfitting to a single beam index. This approach aligns with diverse deployment scenarios and enables more resilient AI/ML evaluation.
3. **Unified AI/ML Beam Management Testing Framework**

**RAN4 should adopt a unified test method that covers both RSRP prediction and beam-ID prediction, with clear fallback and case-specific accuracy rules:*** **RSRP prediction path**
	+ **If the UE reports predicted RSRP values and the prediction accuracy for the top-K beams (K FFS) meets legacy-level measurement accuracy, then only absolute/relative RSRP prediction accuracy requirements apply, with no additional beam ID prediction requirements**.
* **Fallback to beam-ID prediction**
	+ **If RSRP accuracy cannot be guaranteed or if UE reports only beam IDs, testing falls back to beam-ID prediction.**
* **Beam-ID prediction cases**
	+ **Case 1 – Model-Ranked:** **Both beam inclusion and ordering matter. (option 3, within x dB from genie)**
	+ **Case 2 – Probe-Ranked:** **Only beam inclusion within the Top-K set is required.**
* **RAN4 action**
	+ **Define separate accuracy requirements for the two beam-ID cases or select one case and set the corresponding requirements accordingly.**
1. For Case 1 (Model-Ranked), where the UE reports a ranked Top-K list of beam indices directly from the AI/ML model, it is both sufficient and optimal to set K = N in conformance testing. For Case 2 (Probe-Ranked Top-K), we recommend K > N (e.g., K = N+1 or N+2) so the UE can provide a broader beam pool for post-measurement ranking
2. **RAN4 should first determine whether truly scenario-agnostic accuracy requirements are achievable. If not achievable, agree on a scenario list to anchor requirements and conformance tests, starting with BM-Case 1:**
	* **Scenario 1 – Set A ≠ Set B (wide → narrow)**
		+ **Predict SSB beams → CSI-RS beams**
		+ **Inputs:**
			- **(1a) Best-Rx-beam RSRP per Tx (repetition ON)**
			- **(1b) Specific-Rx-beam RSRP per Tx**
	* **Scenario 2 – Set B ⊂ Set A (narrow → narrow )**
		+ **(2-1) Predict all SSB beams from a subset**
			- **(2-1a) Best-Rx input (repetition ON)**
			- **(2-1b) Specific-Rx input**
		+ **(2-2) Predict additional CSI-RS beams from a subset (repetition OFF)**
3. Study on Model Generalization Across Deployment Scenarios for AI/ML Beam Management

**RAN4 is invited to evaluate how different training strategies affect model robustness across diverse deployments:*** **Universal model** – **Train one model on a combined dataset covering multiple deployment scenarios.**
* **Scenario-specific models** – **Train separate stand-alone models for each deployment scenario.**
* **Shared backbone with experts** – **Use a common base model with lightweight scenario-specific expert layers.**

**The study will test all approaches on both Seen (training) and Unseen (new) scenarios to determine whether beam management requires many specialized models or if a shared architecture with limited specialization can meet performance needs, guiding model deployment and standardization in RAN4.** |
| [**R4-2513661**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_116bis/Docs/R4-2513661.zip) | CMCC | ***Proposal 1: For RSRP accuracy for AI/ML based beam management, it is proposed that AI/ML based performance requirement should be no worse than the legacy measurement accuracy requirements.*** ***Proposal 2: based on RAN1 agreements, the legacy L1 report mapping, i.e. L1 part of Table 10.1.6.1-1 and Table 10.1.6.1-2 in TS38.133 can be reused for the report mapping of predicted L1-RSRP, but some wording update is needed.*** |
| [**R4-2514404**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_116bis/Docs/R4-2514404.zip) | Nokia | **Observation 1:** The AoA setups for FR2 RRM test cases (eIFF test system) defined in Annexure A.3.15 of TS 38.133 can have a maximum of 3 active probes during the test.**Observation 2:** The FR2 RRM testing setup reuses the FR2 RF testing setup to reduce the test system complexity and test cost.**Observation 3:** For 2 AoA spherical coverage test, the UE is tested only for one AoA separation as preferred by the UE.**Observation 4:** UE RF test setup to test 2 AoA spherical coverage for multi-Rx chain DL reception was selected to cater the need of different implementations of multi-panel UE, i.e., UEs having 2 antenna panels on opposite, adjacent and same sides of the UE.**Observation 5:** Existing eIFF test system cannot be directly reused for the conformance testing of AI/ML BM use case with CDL based channel model(s).**Observation 6:** 3D MPAC test system is better suited for CDL channel model-based testing.**Proposal 1: RAN4 to agree on selecting the 32 Set A beams, to be emulated during the conformance testing of AI/ML BM use case, from the existing codebook of 128 fixed beams (constructed from a grid of eight elevation angles from –25o to +25 o with ~7.1 o step size and 16 azimuth angles from –60 o to +60 o with 8 o step size) along with the other parameters as defined in Table D.2-1 of TS 38.151.** **Proposal 2: RAN4 to study the impact of selecting different combinations of 32 beams from the 128-beam codebook on the performance with simplified CDL channel model. How to select the 32 beams from 128 beam codebook is currently FFS.****Proposal 3: RAN4 to discuss and explore the feasibility of generating Set A and Set B beams during the conformance testing of the AI/ML-BM use case by selecting unique combinations of available test probes and decide the transmit power of each selected probe. Tx power of each selected probe is FFS.****Proposal 4: RAN4 to discuss and finalize the contents of pre-alignment check which needs to be performed between the TE and DUT before the exact start of the conformance test.****Proposal 5: RAN4 to discuss the feasibility of performing beam sweeping using UE rotation in multiple iterations during the conformance testing of AI/ML based BM use case.****Observation 7:** A spatial filter should be applied to the UMi CDL-C channel model based on the SSB and UE Rx beam, so that only the clusters interacting with them are retained. In this case, the number of clusters will be significantly fewer than the 24 clusters defined in the UMi CDL-C model.**Observation 8:** Based on the reflection and diffusion of the SSB from gNB to UE and its reception by the Rx beam, the power angular spectrum (PAS) represents the “FR2 channel” that needs to be emulated in the anechoic chamber using probes.**Observation 9:** Spatial filter should also be applied to the reduced-cluster channel model according to the beamwidth of SSB and UE Rx beam, so that the interacting clusters can be identified, and the rest of clusters can be ignored.**Observation 10:** Reducing the number of clusters in the CDL model carries a high risk of significantly altering the channel properties, potentially shifting the UMi characteristics to those of UMa or to an undefined channel property.**Proposal 6: It is not recommended to reduce the number of clusters in the UMi CDL-C model for simplification. There is high risk of significantly altering the channel properties, potentially shifting the UMi characteristics to those of UMa or to an undefined channel property.****Proposal 7: A spatial filter should be applied to the CDL-C model first, which can significantly reduce the number of clusters. Next, clusters with insufficient power can be ignored, further reducing the number of clusters. Only the remaining clusters need to be represented in the chamber using probes.****Proposal 8: Endorse the above-mentioned 4-step procedure for representing the clusters in the chamber using probes.****Proposal 9: The reference sensitivity power levels can be reused to determine whether the power of a cluster is too weak and can therefore be ignored.****Proposal 10: No RAN4 requirements for inference need to be defined for the LOS/NLOS indicator agreed by RAN1 for Case 3a.****Proposal 11: Measurement accuracy requirements for gNB Rx-Tx time difference apply for UL SRS-TDCT first path under the condition that the reported measurement is based on N=4 samples.****Proposal 12: Measurement accuracy requirements for UL SRS RSRP apply for UL SRS-TDCP under the condition that the reported measurement is based on N=4 samples.** **Proposal 13: RAN4 to not specify a requirement on training data collection of Part B in Case 3a/3b.** |
| [**R4-2514461**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_116bis/Docs/R4-2514461.zip) | Qualcomm Incorporated | **Observation 1: RAN4 is currently focusing on system level channel, e.g., dense urban macro, to simulate and investigate the performance of spatial only beam prediction.****Observation 2: AI-ML beam prediction requirement defined based on system level channel is more realistic for field deployment.****Observation 3: RAN4 has not calibrated system level channel for FR2 OTA testing****Observation 4: RAN4 has defined multi AoA based RRM testing setup as baseline for AI-ML BM OTA. Companies are invited to bring analysis on what simplified spatial channel model can be emulated in this test setup.****Observation 5: RAN4 has requested companies to bring analysis regarding how CDL channels can be simplified to emulate the channels in the enhanced IFF chamber.****Observation 6: Baseband error at different SNRs can be modelled with Gaussian distribution. For example, at -3 dB SNR, the distribution of absolute baseband measurement error can be fitted with following Gaussian distributions:*** **AWGN: Gaussian with zero mean and sigma = 0.6 dB**
* **TDL-C: Gaussian with mean = -0.12 dB and sigma = 0.85 dB**

**Observation 7: Table 3 contains the results in “narrow” to “narrow” beam prediction scenario.****Table 3: Results of AI-ML based “spatial only” beam prediction (“narrow” to “narrow” scenario)**

|  |  |  |
| --- | --- | --- |
| Metrics | Training and testing without measurement error | Training and testing dataset, along with ground truth for training and testing, with measurement error |
| L1-RSRP absolute accuracy (dB) | +- 3.72 | +- 4.6 |
| Top K/1 without margin (%) | K = 1 | 78.25 | 71.3 |
| K = 3 | 95.6 | 93.6 |
| K = 5 | 98 | 97 |
| Top 1/1 with margin of X dB (%) | X = 1 | 87.45 | 80.53 |
| X = 3 | 94.93 | 91.43 |
| X = 5 | 96.94 | 95.39 |

**Proposal 1: Accuracy of AI-ML BM-case 1 is defined based on the worst-case performance of two test dataset where datasets come from following two scenarios:*** **Dense urban macro**
	+ **Note: This is currently under consideration and captured in the recently agreed simulation assumption of** [**R4-2508081**](http://10.10.10.10/ftp/RAN/RAN4/Inbox/R4-2508081.zip)
* **The CDL channel that gets defined to test AI-ML BM OTA**
 |

## Open issues summary

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

R4-2511890:

**Issue 2-2: Prediction report delay**

Agreement:

* + For aperiodic report, the overall prediction reporting delay at least includes measurement delay +reporting delay
		- The reporting delay includes the inference delay
	+ For semi-persistent and periodic report, the overall prediction reporting delay at least includes measurement delay +inference delay+time for the first available reporting occasion

**Issue 2-3: TCI State Handling**

Agreement:

* + On Detectability and SNR conditions:
		- The UE has sent at least 1 L1-RSRP inference report for the target TCI state 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 ≥ -3 dB
	+ Time conditions:
		- TCI state switch command is received within 1280 ms upon the last transmission of the RS resource for beam predict reporting or measurement
	+ Agreement on QCL relationship
		- If the predicted Tx beam in Set A is QCL Type-D to a known measured Tx beam, where TX beam can be both inside or outside set B, the corresponding Rx beam is known.
		- If the predicted Tx beam in Set A is not QCL Type-D to a known Tx beam, known TCI state conditions shall be based on UE capability.

**Issue 2-9: Metrics/KPIs for beam ID prediction**

Agreement:

* + - If RSRP prediction is reported, absolute/relative RSRP requirements applies to
			* reported beams where SNR side condition is met
			* The exact RSRP reporting including both absolute and relative will be defined in RAN1
			* The related test can only apply to the top-K, where the value of K is FFS, beams.
				+ The legacy beam management test and parameters can be taken as the reference to decide the exact value of K
				+ The feasibility of the test will be further discussed and decided.

**Issue 2-1: Measurement period for inference**

Agreement:

observation period for prediction: (observation period is the amount of time during which UE samples the reference signals to make 1 prediction report)

For case 1 reuse M,N,P from legacy measurement requirements

For case 2, use T\*M, N P (M,N,P same as legacy measurement requirement)

 FFS on T value, T can also be 1. T can also be based on capability

This observation period is to be used for the prediction delay requirement

**Issue 2-4: Activation delay**

Agreement:

* UE has to be ready to start measurements for inference after sending RRC reconfiguration complete
* Inference delay (might include also model loading depending on UE implementation) will be included in the inference reporting delay
* This applies to all reporting schemes

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

Agreement:

 Consider multi AoA based RRM testing setup as baseline(enhanced IFF system with 4 probes used for FR2 RRM conformance testing)

 Single AoA tests can also be performed in the multi AoA based RRM test setup

 Further analyze what spatial channel models can be emulated with this test setup

* + Companies are invited to bring analysis on what simplified spatial channel (multiple clusters coming from multiple directions) model can be emulated in this test setup

 Single AoA tests are not precluded

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

Agreement:

Companies to bring analysis into what simplifications to CDL channels are needed to be able to emulate the channels in the enhanced IFF chamber.

 Take CDL-C Umi as one example

 Other channel models can be further discussed

Open issues from last meeting:

1. Measurement period for inference for case 2 (value of T)
2. Prediction report delay (account for d’)
3. Reporting delay requirements for monitoring (RS-PAI)
4. TCI state handling updates
5. Draft CRs discussion
6. Simulation results and next steps
7. Prediction KPIs
8. Test system setup
9. Test system channel model

### Sub-topic 2-1

*Measurement period for inference – case 2*

The requirements for case 2 are not yet finalized, there are still some FFS from the agreement in RAN4#116

**Issue 2-1: Measurement period for inference – case 2**

Proposals

* + Option 1: T=1
	+ Option 2: T is fixed, different than 1 (2 or 4)?
	+ Option 3: T is based on capability
	+ Option 6: Others
* Recommended WF
	+ Option 1

Proponents of Option 3 should explain why T depends on implementation and what is the system benefit to have different T for different UEs and longer measurement(observation) period

### Sub-topic 2-2

*Reporting delay*

Some agreements on the reporting delay were reached in previous meetings, however , the entire timeline should be clarified so it can be included in the specifications. Some companies are proposing to include d’ and/or Z3’ in the requirements to align with RAN1. This reporting delay will be added to the measurement(observation) period.

**Issue 2-2: Reporting delay**

* Proposals
	+ Option 1: introduce inference/reporting delay as Z3’+d’
	+ Option 2: reference RAN1 specifications, in RAN4 specs mention that UE should report in the reporting occasion occurring after the specified delay
	+ Option 3: no need for an explicit reporting delay, just define a measurement delay which includes the reporting delay
	+ Option 4: others
* Recommended WF
	+ To be discussed

### Sub-topic 2-3

*Reporting delay requirements for monitoring*

Multiple companies proposed to introduce reporting delay requirements for RS-PAI

**Issue 2-3: Reporting delay requirements for monitoring**

* Proposals
	+ Option 1: RAN4 to introduce reporting delay requirements for monitoring
		- Delay to be defined from when UE receives command/configuration to report until the report is sent
	+ Option 2: Do not introduce delay requirements
	+ Option 3: Others
* Recommended WF
	+ Option 1

Further discussion on the exact definition of the delay is needed

### Sub-topic 2-4

*Draft CRs*

Two draft CRs with the agreements so far were submitted, these should be discussed. R4-2513413 and R4-2514311

**Issue 2-4: Draft CRs**

* Points for discussion on the draft CRs
	+ TCI state switching delay
		- Modify the existing clause on known condition or introduce a new one for AI/ML based prediction
		- Adopt the change proposed in R4-2513413 (introduce T0) or modify the text as in R4-2514311
		- Other changes/suggestions
	+ Performance monitoring for AI/ML operation
		- Introduce a new clause for general performance monitoring and subclauses for CSI and Beam prediction monitroting
		- Introduce separate clauses for CSI and beam prediction (not under a general clause for AI/ML monitoring)
		- Introduce subclause under clause 8 or only for measurement reporting delays under clause 9?
		- Others
	+ Functionality activation delay
		- Is this clause needed or not?
		- Should it cover just beam management or also CSI
		- Other comments on title/text
	+ L1-RSRP prediction report requirements
		- Should SSB also be covered or not?
		- Introduce a new clause for prediction(R4-2513413) or modify existing clause on L1-RSRP reporting (R4-2514311)
		- Separate into Measurement reporting requirements and L1-RSRP measurement and prediction requirements?
			* Further separate based on SSB and CSI-RS
		- Other comments on requirements structure, text, etc
	+ Other high level issues for CRs
* Recommended WF
	+ To be discussed

### Sub-topic 2-5

*Simulation results*

**Issue 2-5: Simulation results**

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

### Sub-topic 2-6

*KPIs for prediction*

The latest agreement on the KPIs for prediction were reached in R4-2505105

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

**Agreement:**

* Metrics/KPIs for beam ID prediction, at least for the case if only beam ID is reported:

The successful rate for the correct prediction,

* + The correct prediction is considered as maximum ground-truth RSRP among top-K predicted beams larger than or equal to the ground-truth RSRP of the strongest genie-aided beam(s) – x dB,
	+ FFS on top-N (N</=K) predicted beams have to be considered in the success rate evaluation
		- FFS on N values
	+ FFS on x values

Note: if x=0 and N=1, it means Top-K/1 (%) : the percentage of "the Top-1 strongest beam is one of the Top-K predicted beams"

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

Agreement:

* Absolute RSRP accuracy requirement applies to Top-1 of predicted beams at least
	+ FFS whether to apply to other beams

No further progress has been made on this topic, the main points for debate were whether the requirements should apply for more beams, not just the top predicted and how to ensure the reliability of at least some of the predicted beams.

**Issue 2-6: KPIs for predictionTest system setup**

* Proposals
	+ Option 1: verify half of predicted beam N=K/2+1
	+ Option 2: do not further consider top N (N=1), only consider x based on simulations
	+ Option 3: set x=[1]dB, K=2 or 3
	+ Option 4: Using K = 4, N = 2 and 0=X<2 dB
	+ Option 5: other combination/parameters to ensure that if multiple beams are predicted within a narrow range of absolute RSRP, they are all reliably predicted
		- If there is only a single beam much stronger than others, reliability of prediction of the other beams has no material impact
	+ Option 6: others
* Recommended WF
	+ To be discussed

### Sub-topic 2-7

*Performance requirements for monitoring*

Multiple companies proposed to introduce requirements for RS-PAI reporting.

**Issue 2-7: Test system setup**

* Proposals:
	+ Option 1: introduce accuracy requirements for RS-PAI
	+ Option 2: do not introduce requirements for RS-PAI
* Recommended WF
	+ Option 1

Actual details on how to introduce the requirements and corresponding tests can be discussed in the next meeting

### Sub-topic 2-8

*Test system setup*

A baseline for the test system was agreed in the last meeting, this will have to be further refined

**Issue 2-8: Test system setup**

* Points for discussion
	+ Option 1: support 4 active cross-pol probes in a single plane: 0, 30, 90, 150degrees
		- 
	+ Option 2: emulate UE movement through UE rotation
	+ Option 3: emulate UE movement through changes in power/Tx probe
	+ Option 4: number of beams which can be emulated, is it possible to emulate 32 beams?
* Recommended WF
	+ To be discussed

### Sub-topic 2-9

*Test system channel model*

In the previous meeting it was agreed that a CDL-based channel model will be taken as starting point.Based on the agreements for the baseline test setup in the previous meeting, it should be further discussed which channels can be emulated with the baseline test setup and how to simplify the channel models.

**Issue 2-9: Simplified Channel models**

* Proposals
	+ Option 1: RAN4 to continue evaluation of simplified channel models as proposed in R4-2514248 (R&S – 6 clusters, etc
	+ Option 2: other channel simplification methods:
		- Revmoe the weak clusters that would not influence the UE beam management
		- Merge the clusters that have the same/similer AoAs
		- Reduce values of per-cluster parameters such as CASD, CASA, CZSD, CZSA
	+ Option 3: others
* Recommended WF
	+ To be discussed

It should also be further discuss how to continue the study/evaluation with a simplified channel, how to incorporate this in the simulation and how to check the performance impact.

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

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

## Companies’ contributions summary

|  |  |  |
| --- | --- | --- |
| **T-doc number** | **Company** | **Proposals / Observations** |
| [**R4-2513218**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_116bis/Docs/R4-2513218.zip) | CATT | **Proposal 1: RAN4 to reuse the legacy values and requirements for measurement samples.**  |
| [**R4-2513414**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_116bis/Docs/R4-2513414.zip) | vivo | **Proposal 1: No more new requirement would be needed for case 1, since delay requirement has been agree to take** **framework of the existing reporting delay requirements.****Proposal 2: RAN4 not to define new requirements for positioning case 3. Legacy requirement (if exists) can be reused.****Proposal 3: RAN4 to consider define delay requirements for monitoring Option A-1, A-2 and A-3 for positioning case 1.** |
| [**R4-2513422**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_116bis/Docs/R4-2513422.zip) | Huawei,HiSilicon | **Proposal 1**: For Case 3b, reuse the legacy requirement for reporting of timing information or timing and power information from gNB to LMF.**Proposal 2**: For Case 3a, RAN4 to discuss whether to reuse the legacy requirement for reporting of timing information from 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. |
| [**R4-2513585**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_116bis/Docs/R4-2513585.zip) | Apple | **Proposal for case 3b:** **Proposal 1:** **RAN4 is invited to define the mapping, granularity, quantization, and reporting range for additional samples, and to decide whether phase information should be included and, if so, its representation. Completing these specifications will prevent vendor-specific behaviors, ensure interoperability, and support robust LMF-side model performance** |
| [**R4-2513665**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_116bis/Docs/R4-2513665.zip) | CMCC | ***Proposal 1: for case 1, the number of measurement samples can be 1, 2, or 4 based on the UE capability and LMF request, similar as legacy positioning measurement.*** ***Proposal 2: It is proposed to reuse legacy UL SRS RSRP measurement accuracy requirements for case 3a.*** |
| [**R4-2513670**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_116bis/Docs/R4-2513670.zip) | CMCC | During Rel-19 discussion, following agreements are reached for case 3a.

|  |
| --- |
| **Agreements in RAN4#112bis (R4-2417212)****Issue 3-3: Report applicability for existing reported metrics**Agreement: Extend applicability of existing report mappings for UL-RToA and UL SRS-RSRPP to report UL measurements for AI/ML based positioning use cases 3a/3b.**Agreements in RAN4#114 (R4-2502856)****Issue 3-4: Report mapping for UL SRS-RSRP****Agreement:** * Reuse the report mapping for UL SRS-RSRP to Case 3a

**Agreements in RAN4#114bis (R4-2505105)****Issue 3-3: Rx-Tx time difference for case 3a**Agreement:* Introduce gNB Rx-Tx time difference reporting accuracy requirements
	+ - Reuse legacy accuracy requirements

**Issue 3-4: Report mapping for gNB Rx-Tx time difference**Agreement:* Reuse the existing report mapping for gNB Rx-Tx time difference measurements
 |

Update spec to define requirements for case 3a based on above agreements.Clauses affected 13.1, 13.2, 13.3, 13.6 |
| [**R4-2514093**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_116bis/Docs/R4-2514093.zip) | ZTECorporation,Sanechips | **Observation 1: Nsample defined in legacy positioning equals to 1,2,4 based on UE capability and LMF requests.****Observation 2: UE performers measurement on DL PRS and then the measurement results will be the input of model inference. The measurement behaviour is the same.****Proposal 1: The measurement samples in positioning accuracy enhancement shall be 1,2,4 based on different UE capabilities and LMF requests which same as legacy positioning.** |
| [**R4-2514117**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_116bis/Docs/R4-2514117.zip) | Ericsson | **Observation 1:** Depending on the UE capability, network can configure UE to perform UE based positioning case 1, or UE based positioning case 1 + non-AI/ML based DL-TDoA positioning, or UE based positioning case 1 + non-AI/ML based DL-AoD positioning.**Observation 2:** Depending on UE capability, UE can be configured to perform PRS measurement within measurement gap or outside of the measurement gap in a PRS processing window (PPW) for UE based positioning case 1. **Observation 3:** Depending on the UE capability network can configure UE to perform positioning measurements for UE based positioning case 1 by aggregating PRS resources from 2 or 3 PFLs.**Observation 4**: Following are the scenarios that are valid from the signalling point of view.* Scenario 1: UE-based positioning Case 1
	+ PRS measurement within measurement gap
	+ PRS measurement outside of measurement gap
	+ measurement with and without PRS aggregation from 2 and 3 PFLs
* Scenario 2: Simultaneous UE-based positioning Case 1 and DL-TDoA positioning
	+ PRS measurement within measurement gap
	+ PRS measurement outside of measurement gap

measurement with and without PRS aggregation from 2 and 3 PFLs* Scenario 3: Simultaneous UE-based positioning Case 1 and DL-AoD positioning
	+ PRS measurement within measurement gap
	+ PRS measurement outside of measurement gap
	+ measurement with and without PRS aggregation from 2 and 3 PFLs

**Proposal 1**: RAN4 to define core requirements for the scenario where the UE is only configured to perform UE based positioning case 1.**Proposal 2**: In the core requirement for UE based positioning case 1, Nsample = 1 if UE supports reduced number of samples for PRS measurement, PRS bandwidth is within the active bandwidth part, and magnitude of difference between the serving cell’s SS-RSRP and neighbor cell’s PRS-RSRP is within 6dB. **Proposal 3**: In the core requirement for UE based positioning case 1, Nsample = 2 if UE supports reduced number of samples for PRS measurement, PRS bandwidth is not within the active bandwidth part, and magnitude of difference between the serving cell’s SS-RSRP and neighbor cell’s PRS-RSRP is not within 6dB. **Proposal 4**: In the core requirement for UE based positioning use case 1, Nsample = 4 if UE does not support reduced number of samples for PRS measurement.**Proposal 5**: For UEs that support reduced Rx beam sweeping factor for PRS measurement in FR2, the Rx beam sweeping factor in the core requirement for UE-based positioning case 1 is equal to the value reported by UE as its capability. **Proposal 6**: For UEs that do not support reduced Rx beam sweeping factor for PRS measurement in FR2, default value of 8 is considered in the core requirement for UE-based positioning case 1. **Proposal 7**: Impact of the number of Rx TEGs on the measurement period requirement for UE-based positioning case 1 should not be considered. |
| [**R4-2514352**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_116bis/Docs/R4-2514352.zip) | Nokia | **Case 1**1. For Case 1, RAN4 should define the “Positioning Reporting Requirement” as follows:

Change #1:

|  |
| --- |
| 9.9.X Positioning Reporting Requirements for UE-based direct AI/ML PositioningThe positioning reporting delay is defined as the time between the moment when the periodic measurement report istriggered and the moment when the UE is ready to transmit the positioning report over the air interface. Thisrequirement assumes that the positioning report is not delayed by other LPP signalling on the DCCH. Thispositioning reporting delay excludes a delay uncertainty resulted when inserting the positioning report to the TTI of the uplink DCCH. The delay uncertainty is: 2 x TTIDCCH where TTIDCCH is the duration of subframe or slot or subslot when the positioning report is transmitted on the PUSCH with subframe or slot or subslot duration. This position reporting delay excludes any delay caused by no UL resources for UE to send the measurement report. When receiving a *RequestLocationInformation* message from the LMF via LPP [22, 34], the UE shall be able to send the positioning response, defined in [34], from configured PRS resources for configured TRPs on configured positioning frequency layers, within $T\_{report, total}=T\_{meas}+T\_{inf}+T\_{report, LPP}$,where $T\_{meas}$ is the time taken by the UE to perform the required measurements, $T\_{inf}$ is the time taken to process or infer positioning-related data, and $T\_{report, LPP}$ is the time taken to prepare and transmit the report to the network. |

Change #2:

|  |
| --- |
| 4.5.X Positioning Reporting Requirements for UE-based direct AI/ML PositioningThe positioning reporting delay is defined as the time between the moment when the periodic measurement report is triggered and the moment when the UE is ready to transmit the measurement report over the air interface. The UE will transition to RRC\_CONNECTED state prior to transmitting the positioning report.For positioning performed by a UE in RRC\_IDLE state, the positioning reporting delay excludes all of the following:* additional delay caused other LPP signalling on the DCCH,
* delay uncertainty introduced when inserting the measurement report in the TTI of the uplink DCCH, equal to 2 x TTIDCCH where TTIDCCH is the duration of subframe or slot or subslot when the measurement report is transmitted on the PUSCH with subframe or slot or subslot duration,
* any delay caused by unavailability of UL resources to transmit the measurement report,
* the time needed to transition to RRC\_CONNECTED state to report the measurements.

If a positioning measurement is performed in RRC\_IDLE state, when receiving a *RequestLocationInformation* message from the LMF via LPP [34], the positioning reporting requirements defined in clause 9.9.X is re-used for the positioning reporting delay requirement. |

1. Positioning latency should be lower than the positioning reporting requirement.
2. For Case 1, RAN4 should not define any test case for “Position Reporting Requirement”.
 |
| [**R4-2514353**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_116bis/Docs/R4-2514353.zip) | Nokia | Draft CR - New clauses for positioning reporting delay requirements are introduced. See R4-25143529.9.X Positioning Reporting Requirements for UE-based direct AI/ML Positioning4.5.X Positioning Reporting Requirements for UE-based direct AI/ML Positioning |
| [**R4-2514464**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_116bis/Docs/R4-2514464.zip) | Qualcomm Incorporated | **Observation 1:** Legacy RRM spec does not have any requirement regarding evaluation period of UE’s direct calculation of its co-ordinates.**Observation 2:** Legacy RRM spec has defined measurement period requirements for several UE assisted positioning metrics, e.g.: RSTD, RX-TX time difference, PRS-RSRP, etc.**Observation 3:** UE needs to calculate channel impulse response of different TRPs to run AI-ML based inference and estimate its coordinates.**Observation 4:** UE needs to calculate channel impulse response of different TRPs to calculate RSTD or RX-TX time difference metrics.**Observation 5:** UE needs to calculate RSRP of PRS resources, and not the channel impulse response of TRPs, during PRS-RSRP measurement.**Observation 6:** The eval period of RSTD in each positioning frequency layer has a scaling factor for measurement of same PRS resource with multiple RX TEGs ($k\_{multiTEG,i}$). This scaling factor does not show up in the measurement period of PRS-RSRP.**Observation 7:** UE does not need to transmit to estimate its co-ordinates via AI-ML based inference. On the other hand, the estimation of RX-TX time difference requires UE to receive and transmit.**Observation 8**: The eval period of Rel-17 RSTD (shown in appendix section) assumes the number of PRS RSTD measurement samples to be 1, 2 or 4 depending on UE capability.**Proposal 1:** RAN4 uses the measurement delay of 5G NR RSTD as the baseline of measurement delay of AI-ML based positioning case 1. The number of samples needed for the measurement for positioning case 1 also follows the required number of measurement samples for 5G NR RSTD. |
| [**R4-2514118**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_116bis/Docs/R4-2514118.zip) | Ericsson | **Observation 1**: To ensure that the inferred location information is reported by UE to the network in a timely manner, reporting delay requirement for UE-based positioning case 1 was agreed to be defined in RAN4#116.**Proposal 1**: RAN4 to define two test cases for testing UE-based positioning case 1 reporting delay requirement. One test case for FR1 and another for FR2.**Proposal 2**: 4 cell setup is considered for testing UE-based positioning case 1 reporting delay requirement in FR1 and FR2.**Proposal 3**: To define the test case for UE-based positioning case 1 in FR1, test configurations in Table A.6.6.12.1.1-1, general test parameters in Table A.6.6.12.1.1-2, and cell specific test parameters in Table A.6.6.12.1.1-3, for NR RSTD measurement reporting delay test case for single positioning frequency layer in FR1 SA in the existing specification is reused as a baseline keeping the following aspects in mind:* configuration that are not relevant for UE-based positioning case 1 in Table A.6.6.12.1.1-2 and Table A.6.6.12.1.1-3 shall be removed.
* depending on the signalling designed by RAN2, new parameters (not existing in A.6.6.12.1.1-2 and Table A.6.6.12.1.1-3) may need to be considered.

**Proposal 4**: To define the test case for UE-based positioning case 1 in FR1, test configurations in Table A.7.6.9.1.1-1, general test parameters in Table A.7.6.9.1.1-2, and cell specific test parameters in Table A.7.6.9.1-3, for NR RSTD measurement reporting delay test case for single positioning frequency layer in FR1 SA in the existing specification is reused as a baseline keeping the following aspects in mind:* configuration that are not relevant for UE-based positioning case 1 in Table A.7.6.9.1.1-2 and Table A.7.6.9.1.1-3 shall be removed.
* depending on the signalling designed by RAN2, new parameters (not existing in A.7.6.9.1.1-2 and Table A.7.6.9.1.1-3) may need to be considered.

**Proposal 5**: FR1 testing for UE-based positioning case 1 reporting delay requirement is done under TDL-A propagation condition.**Proposal 6**: FR2 testing for UE-based positioning case 1 reporting delay requirement is done under TDL-C propagation condition. |
| [**R4-2514119**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_116bis/Docs/R4-2514119.zip) | Ericsson | Draft CR* To update list of acronyms in Chapter 3.3 with the expanded forms of UL SRS-TDCT and UL SRS-TDCP measurements.
* To implement report mapping for AI/ML based positioning use case 3a.
* To implement new clauses for report mapping for UL SRS-TDCT and UL SRS-TDCP measurements for AI/ML based positioning use case 3b.
* List of acronyms in Chapter 3.3 is updated with the expanded forms of UL SRS-TDCT and UL SRS-TDCP measurements.
* Changes to existing Chapters 13.1 and 13.2 to implement report mapping for UL-RTOA and gNB Rx-Tx time difference measurements for AI/ML based positioning use case 3a.
* New clauses for report mapping for UL SRS-TDCT and UL SRS-TDCP measurements are introduced AI/ML based positioning use case 3b.
 |
| [**R4-2514404**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_116bis/Docs/R4-2514404.zip) | Nokia | **Proposal 10: No RAN4 requirements for inference need to be defined for the LOS/NLOS indicator agreed by RAN1 for Case 3a.****Proposal 11: Measurement accuracy requirements for gNB Rx-Tx time difference apply for UL SRS-TDCT first path under the condition that the reported measurement is based on N=4 samples.****Proposal 12: Measurement accuracy requirements for UL SRS RSRP apply for UL SRS-TDCP under the condition that the reported measurement is based on N=4 samples.** **Proposal 13: RAN4 to not specify a requirement on training data collection of Part B in Case 3a/3b.** |

## Open issues summary

Agreements from previous meeting in R4-2511890:

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

Agreement:

Take the framework of the existing reporting delay requirements

* Reporting delay includes measurement delay, inference delay and the time needed until the UE send the report
* Check until next meeting on the number of samples needed for the measurements

**Issue 3-3: Report mapping for UL SRS-TDCP measurement**

Agreement:

 Adopt the mapping for UL SRS-TDCP as proposed in R4-2510759

**Issue 3-4: Report mapping for UL SRS-TDCT and UL SRS-TDCP measurements**

Agreement:

Technically endorse CR in R4-2510760, to be later merge with the rest of positioning requirements

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

CR for case 1 requirements

CR for case 3a/3b

Testing for case 1

Requirements for combinations of positioning features or positioning and other features

### Sub-topic 3-1

*CR for Requirements for case 1*

In the previous meeting it was agreed to introduce reporting delay for case 1 based on the legacy requirement. The actual CR should be discussed and agreed.

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

Proposals

* + Option 1: Endorse CR for case 1 in R4-2514353 (Nokia)
	+ Option 2: others
* Recommended WF
	+ Option 1

Please provide any comments on the draft CR

### Sub-topic 3-2

*CR for case 3a/3b*

A draft CR was already technically endorsed in the last meeting. A formal CR could agreed in this meeting for case 3a/3b.

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

Proposals

* + Option 1: merge Clauses 13.3 13.6 from R4-2513670 (CMCC) into R4-2514119 (Ericsson) and endorse final CR
	+ Option 2: Others
* Recommended WF
	+ Option 1

### Sub-topic 3-3

*Testing for case 1*

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

* Proposals
	+ Option 1: Introduce 2 tests, one for FR1 and one for FR”
		- Tests to be based on the legacy tests as proposed in R4-2514118
	+ Option 2: Do not define any tests
	+ Option 3: others
* Recommended WF
	+ To be discussed

### Sub-topic 3-4

*Combinations of positioning features or positioning and other features*

In R4-2514117 it is proposed to introduce requirements for combinations of other features and positioning case 1 or for capabilities introduced for other feature which could have impact on positioning.

**Issue 3-4: Combinations of features and positioning case 1**

* Proposals
	+ Option 1: Introduce multiple requirements for combinations of features and positioning case 1, such examples are
		- UE configured to perform positioning measurements for case-1 and “legacy” PRS based measurements
		- Requirements for the cases in which the UE supports reduce number of samples
		- Requirements for UEs supporting reduced rx beam sweeping factor
	+ Option 2: Do not introduced requirements for such combinations, this will become untractable as there can be too many combinations of features
		- Only keep the baseline requirements
	+ Option 3: others
* Recommended WF
	+ Option 2

# Topic #4: General

This section contains the sub-topics regarding general issues for AI/ML based use cases.

## Companies’ contributions summary

|  |  |  |
| --- | --- | --- |
| **T-doc number** | **Company** | **Proposals / Observations** |
| [**R4-2513410**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_116bis/Docs/R4-2513410.zip) | vivo | **Proposal 1: For Performance Monitoring verification, one way is to design a test to verify** * **Under two sets of different testing conditions, whether UE will notice this change and the changes of the prediction performance will be reflected through the monitoring metric calculated by UE.**

**Proposal 2: For Type 3 performance monitoring for CSI prediction, RAN4 to discuss reporting accuracy requirements for SGCS 1 and SGCS 2.** * **For accuracy metrics for SGCS 1 and SGCS 2, extra non-prediction CSI report would be needed for TE to obtain the ground truth CSI, and CSI (non-predicted).**
* **SGCS 1 is calculated based on predicted CSI for one inference reporting, and ground truth CSI.**
* **SGCS 2 is based on ground truth CSI and CSI (non-predicted) corresponding to the latest CSI-RS transmission occasion not later than CSI reference resource of the inference reporting instance.**

**Proposal 3: For Type 3 performance monitoring for CSI prediction, RAN4 to define reporting delay requirements** **and measurement requirements, which could largely reuse the approach of Beam management monitoring.****Observation 1: For the following scenario mentioned in RAN2, when sending the *RRCreconfigurationcomplete* message, the UE may not be ready to start measurements, which will introduce additional time gap—from the time point when the UE first reports the configuration as inapplicable in the RRCreconfigurationcomplete message to the subsequent time point when the UE reports its applicability to the network via UAI once the configuration becomes applicable.** * **If the UE is not ready for inference by end of RRC processing delay, it reports configuration inapplicable and once the configuration is applicable (i.e. the model is ready to generate inference result based on configuration), UE reports applicability to network via UAI (applicable to all AI/ML based CSI reporting)**

**Proposal 4: RAN4 to define the activation delay to verify how long it takes for the UE to be ready for inference after receiving the RRC Reconfiguration message. The starting point and ending point of this delay are as follows:** * **Upon the reception of the RRC Reconfiguration message to the UE reporting that the configuration is applicable (via *RRCReconfigurationComplete* or via UAI).**
 |
| [**R4-2513419**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_116bis/Docs/R4-2513419.zip) | Huawei,HiSilicon | **Observation 1:** Different delay requirements are identified for different performance monitoring reporting types.**Proposal 1:** RAN4 defines different delay requirements for periodic, aperiodic and semi-persistent monitoring reporting. **Observation 2:** NW controls when the uplink transmission resource for AI-based reporting is not available.**Proposal 2:** RAN4 not to define deactivation requirements. **Proposal 3:** If a mixed dataset is created for testing generalization, the mixed dataset should be static. |
| [**R4-2513582**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_116bis/Docs/R4-2513582.zip) | Apple | 1. Option 1 (pre-activation testing) ensures assurance but can be slow, costly, and impractical for frequent updates. Option 2 (post-activation LCM monitoring) is scalable but risks false triggers, overhead, and delayed detection if not carefully designed
2. **Pre-activation protects the network from immediate catastrophic failures at the moment of deployment**, while **post-activation protects against gradual or unforeseen degradation during the model’s operational life**. Both are needed in the hybrid approach to cover different risks and time horizons
3. Shadow inference is essential for Option 2 because post-activation monitoring only gathers KPIs from the model actively serving inference. Without shadow mode, standby models remain untested in live conditions, risking undetected drift or degradation until swapped
4. Hybrid LCM for Post-Deployment AI/ML Models

Adopt a hybrid life-cycle management (LCM) framework: **pre-activation validation for major updates** and **post-activation monitoring for all updates**  * **Major updates (architecture change, new RF front-end, large dataset shift)**
	+ **Require pre-activation validation in lab or on-device test mode.**
	+ **Establish or refresh a KPI envelope as baseline.**
* **Minor updates (small fine-tune, parameter tweak)**
	+ **May skip pre-activation.**
	+ **Must retain prior KPI envelope and enter direct monitoring.**
* **Post-activation monitoring (all updates)**
	+ **Continuous KPI checks under RAN4/RAN1 metrics (e.g., Top-M/Top-K beam accuracy).**
	+ **Prefer shadow inference to track standby models.**
* **Rationale**
	+ **Pre-activation blocks immediate large failures; post-activation catches drift and long-term degradation.**
1. Multi-Model LCM and Shadow Inference

**Investigate require shadow inference for all non-active models in multi-model deployments.Every standby model must process the same live input stream as the active model, with its KPIs checked against genie references or active-model baselines.*** **Applicability: Applies to any UE storing multiple AI/ML models under Option 2 post-activation monitoring.**
* **Functionality: Standby models run in parallel “shadow” mode, producing KPIs (e.g., beam accuracy, RSRP error) that never influence network decisions.**
* **Objective: Detect drift, dataset mismatch, or hidden faults in standby models before activation, ensuring they remain deployment-ready.**
* **Assurance: Guarantees that every stored model is continuously validated, not just the currently active one. Maintains service quality during frequent updates and dynamic model switching**
1. Standardized Criteria for Major vs. Minor Model

**RAN4 shall specify clear rules to classify AI/ML model updates as major or minor, determining whether pre-activation validation is required while ensuring all models undergo post-activation LCM monitoring.*** **Major update triggers: full retraining, architecture change, new channel models or RF front-end profiles, significant dataset distribution shift, or inference-pipeline modification.**
* **Minor update triggers: limited fine-tuning, small parameter or dataset refresh without altering statistical characteristics.**
* **Purpose: provide consistent, enforceable criteria so vendors and operators know when lab or on-device pre-activation testing is mandatory.**
* **Assurance: guarantees that every update, regardless of classification, enters continuous post-activation monitoring to detect long-term drift.**
1. Scenario- and Model-Specific KPI Envelopes

**RAN4 shall require that every AI/ML model carries a scenario- and model-specific KPI envelope for post-deployment monitoring, providing the correct performance baseline for its intended environment.*** **Scenario-specific: KPI thresholds reflect the radio conditions (e.g., urban macro, rural macro, high-Doppler) the model targets.**
* **Model-specific: Each model or adapter version maintains its own KPI envelope aligned with its architecture and training data.**
* **Major updates: Generate a new KPI envelope from pre-activation validation results.**
* **Minor updates: Inherit the validated KPI envelope from the prior deployment.**
* **Purpose: Ensure accurate and reliable live performance tracking, avoiding false alarms or missed degradations.**
 |
| [**R4-2513664**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_116bis/Docs/R4-2513664.zip) | CMCC | ***Proposal 1: it is proposed that the agreements on activation for beam prediction is applied for CSI prediction. In detail, for CSI prediction, UE has to be ready to start measurements for inference after sending RRC reconfiguration complete.*** ***Proposal 2: for generalization test, it is proposed to consider >1 test per UE capability if the granularity of UE capability is rough.***  |
| [**R4-2513689**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_116bis/Docs/R4-2513689.zip) | Korea Testing Laboratory | **Observation 1: If the first report under the updated model arrives before a network-known decision-time effective-from boundary, it cannot be deterministically placed on either side of the transition. The ensuing RS-PAI window may mix pre/post monitoring occasions and make pass/fail scheduler-phase dependent.****Observation 2: Without an explicit effective-from boundary, the network cannot reliably separate pre-vs post-transition window; additionally, since RS-PAI conveys only** $ N\_{p}$ **(not the used link-offset k), it cannot audit which occasions were linked to which inference reports.****Observation 3: BM-Case-2 requires at least support to report one configured time instance; under slot-based linking (minimal offset, X=64), this single-instance view cannot disambiguate pre- vs post-transition reports across prediction horizons****Observation 4: In BM-Case-2, the earliest** $f$**-instance is tied to the last Set-B occasion (no later than the reference) and Set-B is fixed per report; therefore time instances cannot be aligned to the transition moment; hence pre-/post-transition reports remain indistinguishable without an explicit decision-time boundary.****Observation 5: Absent a network-known decision-time effective-from boundary, identical updates produce different RS-PAI windows across deployments, yielding divergent** $ N\_{p}$ **and conflicting pass/fail (good/moderate/bad) outcomes under the same Option-2 configuration.****Observation 6: Model generalization and measurement fairness are orthogonal. Even if a model is well-generalized, changes to the monitoring set (subset bitmap or 1:1) alter the distribution of monitoring occasions; therefore, an effective-from-aligned switch-over and decision-time gating are required so that RS-PAI windows for performance monitoring report never mix pre-/post-change monitoring occasions.****Observation 7: To preserve comparability across deployments, align any fine-tuning or model update to a single decision-time effective-from boundary. If the monitor-to-Set-A mapping changes (subset bitmap or 1:1), perform an effective-from-aligned switch-over at the effective-from time so that no RS-PAI window spans two different mappings.****Observation 8: In BM-Case-2 of UE-side model, validation must adhere to the RAN1 decision-time structure with slot-based time instances, an explicit time instance indicator, and a fixed Set B per report, so instances must not be mixed or reinterpreted by delivery time variations.****Observation 9: Because monitoring reuses the legacy priority rule** $Pri\_{iCSI}\left(y,k,c,s\right)$ **with k=0, delivery-time variations (scheduler jitter, CPU load) must not be used to reclassify reports across the decision-time boundary; validation remains on the decision-time axis.****Observation 10: When the monitor-to-Set-A mapping (subset bitmap or 1:1) changes, perform an effective-from-aligned switch-over at the effective-from boundary. If only the model behaviour changes, keep the mapping and use an effective-from boundary so that RS-PAI windows do not mix pre-/post-change occasions.****Observation 11: Associated ID aligns what is used across training and inference, but it does not convey when a new behaviour becomes active; it cannot replace a decision time effective from boundary.****Observation 12: In BM-Case-2, the valid P/SP/AP inference-and-monitoring combinations are fixed, so report type pairing alone cannot distinguish pre- vs post-transition reports; a decision-time effective-from boundary is still required.****Proposal 1: Discuss adopting a minimal, explicit “effective-from” UE uplink indication that marks when the updated behavior applies and identifies the relevant functionality/anchor group (per RAN1-agreed use cases).** **Proposal 2: Discuss minimal delivery and classification rules on the decision-time axis to ensure deterministic RS-PAI boundaries. Each RS-PAI window for a performance-monitoring report is classified solely by the effective-from boundary, and delivery-time variations shall not trigger reclassification.****Proposal 3: When the monitor-to-Set-A mapping changes (subset bitmap or 1:1), use an effective-from-aligned switch-over so that no RS-PAI window spans two mappings.** |
| [**R4-2513953**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_116bis/Docs/R4-2513953.zip) | CAICT. | ***Proposal 1: Use the proposed table as starting point for Rel-19 NR\_AIML\_air UE feature discussion.***

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Features** | **Index** | **Feature group** | **Components** | **Prerequisite feature groups** | **Need for the gNB to know if the feature is supported** | **Applicable to the capability signalling exchange between UEs (V2X WI only)”.** | **Consequence if the feature is not supported by the UE** | **Type****(the ‘type’ definition from UE features should be based on the granularity of 1) Per UE or 2) Per Band or 3) Per BC or 4) Per FS or 5) Per FSPC)** | **Need of FDD/TDD differentiation** | **Need of FR1/FR2 differentiation** | **Capability interpretation for mixture of FDD/TDD and/or FR1/FR2** | **Note** | **Mandatory/Optional** |
| 59. NR\_AIML\_air | [59-1-x] | UE-side beam prediction for BM Case1 | [Knows RX beam of predicted TX beam in setA in BM-case 1 if the predicted Tx beam is not QCL Type-D to a known TX beam] | **Pre-requisitng FG: 58-1-2 (RAN1 feature for BM-case-1)** | yes | N/A | [In BM-case 1, network needs to transmit additional samples of reference signal corresponding to the predicted TX beam of setA so that UE can sweep its RX beams during TCI state switch process when the RS resource for predicted Tx beam is the RS in target TCI state or QCLed to target TCI state.] | [Per Band] | [TDD] | [FR2-1 only] | N/A |  | [Optional with capability signaling] |
|  | [59-1-y] | UE-side beam prediction for BM Case2 | [Knows RX beam of predicted TX beam in setA in BM-case-2 if the predicted Tx beam is not QCL Type-D to a known TX beam] | **Pre-requisitng FG: 58-1-4 (RAN1 feature for BM-case-2)** | yes | N/A | [In BM-case 2, network needs to transmit additional samples of reference signal corresponding to the predicted TX beam of setA so that UE can sweep its RX beams during TCI state switch process when the RS resource for predicted Tx beam is the RS in target TCI state or QCLed to target TCI state.] | [Per Band] | [TDD] | [FR2-1 only] | N/A |  | [Optional with capability signaling] |

 |
| [**R4-2514307**](https://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_116bis/Docs/R4-2514307.zip) | Nokia | **Proposal 1: RAN4 needs to define accuracy requirements for UE-assisted monitoring where the UE reports the beam prediction accuracy, RS-PAI, in BM-Case1 and BM-Case2. The accuracy of RS-PAI should be tested at least in static radio conditions.****Observation 1:** RAN4 can define reporting accuracy requirements for AI/ML-based CSI prediction by using a test framework that monitors SGCS consistency under static radio conditions.**Proposal 2: RAN4 should define reporting accuracy requirements for CSI prediction performance monitoring using a test framework that evaluates the stability of reported values in a fixed environment.****Observation 2:** Less generalized functionalities across a set of scenarios can result in frequent switching of functionality/configurations resulting in performance degradation.**Proposal 3: RAN4 to use the following metric, indicative of generalization capabilities of AI/ML functionality/configurations, to verify the generalization performance of the functionality/configuration in different scenarios:**$$Overall Generalization Score= \frac{(Verdict\_{Scenario\_{1}}+…+Verdict\_{Scenario\_{N}})}{N}×100\%$$* **Where N is the number of selected scenarios for generalization testing**
	+ **The value of N to be discussed**
* **Verdict\_j = 1 if the performance requirements are fulfilled for scenario *j***
* **Verdict\_j = 0 otherwise**

**Proposal 4: RAN4 should consider discussing on a generalization test framework that can help to evaluate the generalization capabilities of the AIML enabled functionalities.** |

## Open issues summary

Many of the issues raised in the general agenda are treated in the respective use cases, only a few separate ones are listed below.

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

1. UE features for AI/ML
2. Generalization

### Sub-topic 4-1

*UE features*

In previous meetings it was agreed to introduce the UE features below.

**Issue 4-1: UE Features**

Proposals

* + Option 1: Confirm UE feature as in the table below

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Features** | **Index** | **Feature group** | **Components** | **Prerequisite feature groups** | **Need for the gNB to know if the feature is supported** | **Applicable to the capability signalling exchange between UEs (V2X WI only)”.** | **Consequence if the feature is not supported by the UE** | **Type****(the ‘type’ definition from UE features should be based on the granularity of 1) Per UE or 2) Per Band or 3) Per BC or 4) Per FS or 5) Per FSPC)** | **Need of FDD/TDD differentiation** | **Need of FR1/FR2 differentiation** | **Capability interpretation for mixture of FDD/TDD and/or FR1/FR2** | **Note** | **Mandatory/Optional** |
| 59. NR\_AIML\_air | [59-1-x] | UE-side beam prediction for BM Case1 | [Knows RX beam of predicted TX beam in setA in BM-case 1 if the predicted Tx beam is not QCL Type-D to a known TX beam] | **Pre-requisitng FG: 58-1-2 (RAN1 feature for BM-case-1)** | yes | N/A | [In BM-case 1, network needs to transmit additional samples of reference signal corresponding to the predicted TX beam of setA so that UE can sweep its RX beams during TCI state switch process when the RS resource for predicted Tx beam is the RS in target TCI state or QCLed to target TCI state.] | [Per Band] | [TDD] | [FR2-1 only] | N/A |  | [Optional with capability signaling] |
|  | [59-1-y] | UE-side beam prediction for BM Case2 | [Knows RX beam of predicted TX beam in setA in BM-case-2 if the predicted Tx beam is not QCL Type-D to a known TX beam] | **Pre-requisitng FG: 58-1-4 (RAN1 feature for BM-case-2)** | yes | N/A | [In BM-case 2, network needs to transmit additional samples of reference signal corresponding to the predicted TX beam of setA so that UE can sweep its RX beams during TCI state switch process when the RS resource for predicted Tx beam is the RS in target TCI state or QCLed to target TCI state.] | [Per Band] | [TDD] | [FR2-1 only] | N/A |  | [Optional with capability signaling] |

* + Option 2: other features are needed
* Recommended WF
	+ Option 1

### Sub-topic 4-2

*Generalization and monitoring*

**Issue 4-2: Generalization and monitoring**

Proposals

* + Option 1: Introduce a generalization capability framework based on functionality/configuration for different scenarios

**RAN4 to use the following metric, indicative of generalization capabilities of AI/ML functionality/configurations, to verify the generalization performance of the functionality/configuration in different scenarios:**

$$Overall Generalization Score= \frac{(Verdict\_{Scenario\_{1}}+…+Verdict\_{Scenario\_{N}})}{N}×100\%$$

* **Where N is the number of selected scenarios for generalization testing**
	+ **The value of N to be discussed**
* **Verdict\_j = 1 if the performance requirements are fulfilled for scenario *j***
* **Verdict\_j = 0 otherwise**
	+ Option 2: this is not needed and difficult to introduce, how to define the verdicts and scenarios, how will it help the system if the score is low in some scenarios and high in others?
	+ Option 3: others.
* Recommended WF
	+ Option 2