**3GPP TSG RAN WG1 #114 R1-23xxxxx**

**Toulouse, France, August 21st – August 25th, 2023**

Source: Moderator (OPPO)

Title: Summary#5 for other aspects on AI/ML for beam management

Agenda Item: 9.2.3.2

Document for: Discussion and Decision

# Introduction

The Rel-18 WID of AI/ML for NR Air Interface focuses on a subset of three typical use cases:

1. CSI feedback enhancement
2. Beam management
3. Positioning accuracy improvement.

This document focuses on the other aspects of AI/ML for beam managements, including representative sub use cases and potential specification impact.

Regarding the file names, companies are encouraged to follow the guidance of R1-2203012 (Page 16) as below:

|  |
| --- |
| * + - To avoid ending-up with too long file names and downloading/opening issues, the following naming convention is recommended:       * Keep the previous company’s name (only the most recent one) in the filename, e.g.         + 5/Summary-1-v000-Moderator (HW)         + 5/Summary-1-v001-LG         + 5/Summary-1-v002-LG-CATT         + 5/Summary-1-v003-CATT-vivo         + 5/Summary-1-v004-Moderator(HW)       * It helps identifying on which previous version your input is based on and solve any crossing emails issue. Note the use of 3digit version numbers in the file names. |

In the following sections, the company proposals are summarized, and offline proposals are drafted based on company contributions for discussion/input.

# Spec impact of Data collection

## General/common aspects

In previous RAN1 meeting(s), the agreement(s)/conclusion(s) were made as below:

|  |
| --- |
| **RAN1#110**  Agreement  For the data collection for AI/ML model training (if supported), study the following aspects as a starting point for potential necessary specification impact:   * Signaling/configuration/measurement/report for data collection, e.g., signaling aspects related to assistance information (if supported), Reference signals * Content/type of the collected data * Other aspect(s) is not precluded   **RAN1#113**  Agreement (AI 9.2.1)  Consider at least the following aspects and if applicable, the corresponding potential specification impact related to data collection:   * Measurement configuration and reporting * Contents, type and format of data including:   + Data related to model input   + Data related to ground truth   + Quality of the data   + Other information * Signaling of assistance information for categorizing the data   + Note: The study should consider the feasibility of disclosure of proprietary information * Signaling for data collection procedure * Note 1: Use-case specific details can be studied in respective agenda items * Note 2: Signaling mechanism details can be studied by appropriate working groups. |

The related proposals in tdocs are copied as below:

|  |  |
| --- | --- |
| NVIDIA[3] | *Proposal 5: For AI/ML model training for beam prediction in spatial/time domain, study potential specification impact related to training data type, training data source determination, and assistance signalling and procedure for training data collection.* |
| Intel[9] | *Observation 3: Training dataset construction using 3GPP specified measurement and reporting framework may be advantageous for harmonizing deployment of proprietary AI/ML models.* |
| Apple[20] | *Proposal 3: for discussion on AI/ML BM, the purpose for “data collection” should always be clear, e.g., data collection for training data, UE reporting for inference, UE reporting for performance monitoring.* |
|  |  |

|  |  |
| --- | --- |
| Company | Comments |
|  |  |
|  |  |
|  |  |
|  |  |

## Network-side AI model

In previous RAN1 meeting(s), the agreement(s)/conclusion(s) were made as below:

|  |
| --- |
| **RAN1#112bis-e**  Agreement  Regarding data collection for NW-side AI/ML model, study the following options (including the combination of options) for the contents of collected data,   * Opt.1: M1 L1-RSRPs (corresponding to M1 beams) with the indication of beams (beam pairs) based on the measurement corresponding to a beam set, where M1 can be larger than 4, if applicable   + FFS: the range of M1 * Opt.2: M2 L1-RSRPs (corresponding to M2 beams) based on the measurement corresponding to a beam set, where M2 can be larger than 4, if applicable   + FFS: the range of M2 * Opt.3: M3 beam (beam pair) indices based on the measurement corresponding to a beam set, where M3 can be larger than 4, if applicable   + FFS: the range of M3 * FFS: How to select the M1/M2/M3 beam(s) or beam pair(s) * Note: Overhead, UE complexity and power consumption should be considered for the above options   Agreement  Regarding data collection for NW-side AI/ML model, study necessity, benefits and beam-management-specific potential specification impact from RAN1 point of view on the following additional aspects   * Mechanism related to the reporting * Additional information for content of the reporting   + FFS: Information associated with or configured for the reported data samples, e.g., timestamps, SNR, data quality, etc. * Reporting overhead reduction * Note1: non-3GPP based solution is a separate issue. * Note2: The framework corresponding to higher layer(s) are up to the associated WG(s) * Note 3: Overhead, UE complexity and power consumption should be considered |

The related proposals in tdocs are copied as below:

|  |  |
| --- | --- |
| H3C[1] | *Proposal5：Regarding data collection for NW-side AI/ML model，the rang of M can be set as follow*  *• Opt1 can be used for model training and monitoring, M1 can be up to 64*  *• Opt2 can be used for model training and inference, M2 can be up to 64*  *• Opt3 can be used for model monitoring, M3 can be up to 8*  *Proposal6: For the NW-side model, the NW needs to instruct the terminal to adopt which mode for training data reporting. If dynamic omission of some data is required, we recommend using high-layer signaling for reporting. (data collection for training)* |
| FUTUREWEI[2] | *Proposal 1: Accept the following FL proposal on overhead reduction of data collection.*  *Regarding data collection for NW-side AI/ML model of BM-Case1 and BM-Case2, study the potential specification impact of overhead reduction from the following aspects:*   * *Whether/how ~~omit some data~~ to be selective on data to be reported.* * *Whether/how to compress the reported content (e.g. lager quantization step)*   *Note: For the different purposes of data collection, the overhead reduction mechanisms and corresponding specification impacts may be different.*  *Proposal 2: Regarding data collection for NW-side AI/ML model of BM-Case1 and BM-Case2, study necessity, and potential specification impact on the following aspect*   * *Mechanism to ensure the common understanding between NW and UE on Rx beam assumption for reported measurement result(s)* |
| Huawei[4] | *Observation 6: Considering the low frequency of training/monitoring data collection and the large number of UEs to contribute the data samples, the average air-interface overhead for per UE is negligible.*  *Proposal 8: RAN1 to further study the potential spec impact of data collection for training and monitoring to increase the number of reported RSRPs, SSBRIs and CRIs beyond 4.*  *Proposal 9: For data collection for a network-side AI/ML model, the legacy beam sweeping and CSI report mechanism can be reused as a starting point.*  *Proposal 10: Regarding data collection options for network-side AI/ML model, consider Opt.1(M1 L1-RSRPs and indication) with higher priority than Opt.2(M2 L1-RSRPs) due to smaller overhead.*  *Observation 7: For the container of the reported data samples in data collection, L1 signaling has lower latency and is applicable to training, inference and monitoring, while RRC signaling has larger latency and is applicable only to training.*  *Proposal 11: For training/monitoring data collection, both L1 signaling and RRC signaling can be considered.*  *• Note: The legacy signaling framework (e.g., CSI report for L1 and MDT signaling for RRC) can be reused.* |
| Spreadtrum[5] | *Proposal 4：For data collection at NW-side, option 1 and option 2 should be considered with high priority.*  *Proposal 5：For DL beam pair prediction, the implicit mechanism to ensure the common understanding between NW and UE on Rx beam assumption should be considered.*  *Proposal 6：For DL Tx beam prediction, the mechanism to ensure the common understanding between NW and UE on Rx beam assumption for reported measurement result(s) is not needed.* |
| IDC[6] | *Observation 17: The current NR specification supporting UE reporting with up to 4 best CRIs/SSBRIs with L1-RSRP or L1-SINR can be very limited for a network-side AI/ML model.*  *Proposal 28: Consider increasing number of CRIs/SSBRIs (e.g., 8 CRIs/SSBRIs).*  *Proposal 29: Support both Option 1 and Option 2 for a NW side AI/ML model where the selection between Option 1 and Option 2 could be based on types of Set B and/or AI/ML model inputs.*  *• Option 1: M1 L1-RSRPs (corresponding to M1 beams) with the indication of beams (beam pairs) based on the measurement corresponding to a beam set, where M1 can be larger than 4, if applicable.*  *• Option 2: M2 L1-RSRPs (corresponding to M2 beams) based on the measurement corresponding to a beam set, where M2 can be larger than 4, if applicable.*  *Proposal 30: Support a range of M1 to be 12.5% to 50% of the size of Set A as a baseline. The range of M2 should be equal to or greater than M1.*  *Proposal 31: Enhancement of ANR procedure to exchange AIML-related information between multiple cells should be studied.* |
| Vivo[7] | *Proposal 18: Regarding the data collection for AI/ML model training at NW side, study potential specification impact on resource configuration:*  *• Specific beam pair resource configuration for Set A*  *• Enhanced P3+P2 resource configuration that Rx beam assumption of P2 resource measurement is the best Rx beam searched from P3 procedure for performance improvement*  *Proposal 19: Regarding the data collection for AI/ML model training at NW side, study potential specification impact on assistance information:*  *• Proprietary processed Rx beam information as assistance information from UE to NW, including measured Rx beam information, expected Rx beam information, and best Rx beam information.*  *Proposal 20: Regarding the data collection for AI/ML model training at NW side, study potential specification impact on measurement report:*  *• UE measures the beams of Set A and reports M1 L1-RSRPs optionally with M2 RS indicators, where M1 and M2 can be larger than 4.*  *- If M1 is equal to the number of beams or beam pairs in Set A (noted as X), corresponding RS indicators may be not needed.*  *- If M1 is smaller than X/2, corresponding M2 RS indicators are needed*  *- If M1 is smaller than X, but larger than X/2, RS indicators are needed for indicating M2 beams or beam pairs in Set A not included in the measurement report.*  *Proposal 21: Regarding the data collection for AI/ML model training at NW side, study potential specification impact on report overhead reduction:*  *• Reducing unnecessary L1-RSRP report where the omitted L1-RSRPs may be lower than a pre-defined threshold*  *Proposal 22: Regarding the data collection for AI/ML model training at NW side, study potential specification impact on quantization enhancement for RSRP quality improvement:*  *• High-precision L1-RSRP quantization*  *• Multi-resolution L1-RSRP quantization, e.g. high-resolution quantization for a group of best RSRPs and low-resolution quantization for others* |
| ZTE[8] | *Observation 1: The content of collected data for Set A can be Top-1/K beam(s) with/without the corresponding L1-RSRPs or L1-RSRPs of all the beams in Set A.*  *Observation 2: The content of collected data for Set B can be all or partial measurement results in the measured beam set (i.e., Set C), where the partial beams to be reported can be the Top-K beams with larger measured L1-RSRP or the beams whose measured L1-RSRPs exceed a pre-defined threshold.*  *Proposal 2: For the measurement report of the measured beam set (i.e., Set C), study enhanced reporting of variable number of Top-K beams based on a pre-defined threshold.*  *Observation 3: One or two RS resource sets for beam measurement can be configured to the UE for data collection of NW-side AI/ML model, which are dependent on the beam set construction of Set A and Set B.*  *Proposal 3: Support to study enhancement on resource configuration aspects for data collection.*  *Proposal 4: For data collection from UE to NW side, support to study explicit or implicit Rx beam ID reporting method, especially for beam pair prediction.*  *Proposal 5: If all measurement results of Set A/B need to be reported to gNB, study enhanced reporting methods for the purpose of reporting overhead reduction, e.g., beam ID can be obtained implicitly from the reporting order of all measured RSRPs.*  *Proposal 6: The Rx beam assumption for model training/inference can be aligned between NW and UE by the QCL indication/configuration and no additional spec impact is foreseen.*  *Observation 4: Due to the relatively long data collection duration, the wireless link between gNB and UE may change rapidly, resulting in unguaranteed data sample quality.*  *Proposal 7: Regarding data collection for NW-side AI/ML model of BM-Case1 and BM-Case2, study the potential specification impact of overhead reduction from the following aspects:*  *• Data omission (e.g., according to data quality)*  *• Compression on reported content (e.g., lager quantization step, threshold-based measurement report, measurement report with a pre-defined order)* |
| Intel[9] | *Proposal 3: For data collection for NW-side AI/ML model, support only L1-RSRP reporting. Reporting of only beam indices does not need to be supported.* |
| Ericsson[11] | *Proposal 6 Regarding the FFS on M1/M2/M3 for training data collection mechanism with a NW-side AI/ML model trained at NW side:*  *- M1/M2/M3 is equal to the size of the beam set*  *Proposal 7 Regarding data collection for NW-side AI/ML model of BM-Case1 and BM-Case2, study the potential specification impact of overhead reduction from the following aspects:*  *• Whether/how to omit some data (e.g., avoid report of duplicated/similar samples)*  *• Whether/how to compress the reported content (e.g., lager quantization step or time-domain compression)*  *• Note: For the different purposes of data collection, the overhead reduction mechanisms and corresponding specification impacts may be different.* |
| Google[12] | *Proposal 17: For AI/ML in gNB side, study the following potential enhancement to reduce the L1-RSRP measurement and quantization error.*  *• CSI-RS coverage enhancement*  *• More advanced receiver to reduce measurement error*  *• High-resolution quantization scheme to reduce quantization error*  *Proposal 19: Support to configure two CMR sets in a CSI report configuration for data collection, where the first CMR set is for the report for model input and the second CMR set is for the report for model output.*  *Proposal 20: Support to maintain the same understanding between the NW and UE on when to perform the measurement for data collection*  *• UE does not need to receive another downlink signal in the same symbol when it performs the measurement for the SSB/CSI-RS for data collection*  *Proposal 21: Support to report a hypothetical measurement error for each reported beam for data collection.* |
| LG[13] | *Proposal #8: RAN1 should focus on potential enhancement on L1 beam report, and leave the higher-layer based approach for beam measurement collection to RAN2.* |
| CATT[14] | *Proposal 1: For DL beam pair prediction with a NW-side model, considering the generalization of different UE Rx beam shapes/patterns, study how to report/send relative Rx beam information when preserving sensitive proprietary information and considering the reporting overhead and storage overhead.*  *Proposal 6: Regarding data collection for NW-side model, study the following options for reporting overhead reduction:*  *• Option 1: Reducing the number of reported beams (beam pairs)*  *• Option 2: Reducing the number of reported index*  *• Option 3:* *Reducing the number of bits for RSRP quantization.* |
| NEC[15] | *Proposal 4: Regarding data collection for BM-Case1/2 with NW-side AI/ML model, study to distinguish between input data and output data in the beam report for data collection.*  *Proposal 5: Regarding data collection for BM-Case1/2 with NW-side AI/ML model, study to omit the duplicate data between input data and output data to avoid unnecessary reporting overhead.*  *Proposal 6: Regarding data collection for BM-Case1/2 with NW-side AI/ML model, study to report the indicator of the selected beam pattern to NW if the beam pattern of Set B is variable.* |
| Fujitsu[16] | *Proposal 1: Regarding the data collection for NW-side AI/ML model, study the potential specification impacts on the UE behavior of beam reporting.*  *Proposal 2: Study the potential specification impacts on the enhanced signaling/procedure of reporting configuration of data collection for NW-side AI/ML model.* |
| CMCC[17] | *Proposal 2: For data collection of beam pair prediction at NW side, study how to align the understanding of beam pair ID between the NW and the UE.*  *Proposal 3: For data collection of DL Tx beam prediction at NW side, the Rx beam assumption should be aligned between the network and UE.*  *Proposal 4: Regarding data collection for NW-side model, study the enhancement of L1 reporting of more than 4 beams/beam pairs.*  *Proposal 5: Regarding data collection for NW-side model, study the potential specification impact of overhead reduction from the following aspects:*  *· Whether/how to omit some data (e.g. RSRP threshold)*  *· Whether/how to compress the reported content (e.g. larger quantization step)* |
| Nokia[19] | *Proposal 22. On selecting the content of data collection for NW side model performance monitoring, support both Opt1 and Opt3.*  *Proposal 23. For NW-sided data collection, RAN1 to consider unnecessary overhead on measurement reporting coming from quantized L1-RSRPs values very close to noise level.*  *• Reduce the measured quantized reporting overhead*  *Observation 8. Common understanding between NW and UE on Rx beam assumption for reported measurement result(s) for Tx-Rx beam pair prediction is required due to no explicit information about the Rx beam used to measure the Tx-Rx beam pairs that are communicated by the UE to the NW on one hand, and measurement inaccuracies on the reported data on the other hand.*  *Proposal 24. RAN1 to support the potential specification impact on the following aspect.*  *• Mechanism to ensure the common understanding between NW and UE on Rx beam assumption for reported measurement result(s)* |
| Apple[20] | *Proposal 1a:*  *• For Model training at the NW side & inference at the NW side (Alt. 1), study efficient signalling of set B selection or beam selection and RSRP representation.*  *• For Model training at the UE side & inference at the UE side (Alt. 2), study model generalization performance.*  *Observation 3:*  *• Alt. 1 (NW side training/NW side inference) does not require disclosure of network implementation information.*  *Proposal 2: RAN1 should prioritize the study on quantization error’s impact to AI/ML model inference performance over measurement error.*  *Proposal 4: the correlation among RSRPs for the same Tx beams at Occasions for BM Case-2, and achieve reduced feedback overhead by exploiting the temporal correlation in their RSRPs.* |
| Xiaomi[21] | *Proposal 16: For data collection of NW-side AI/ML model, support to define a time window for each report to include more than one data sample and configure a number of report to stop the data collection.* |
| DCM[22] | *Observation 5: Study the potential specification impacts on the reported measurement selection policy, e.g., filtering the reported measurements whose value is larger than certain threshold.* |
| OPPO[23] | *Observation 5: For the simplest case of BM-Case1 and BM-Case2 with NW-side AI/ML model trained at NW, the model inputs (via enhanced beam reporting) and labels (via legacy beam reporting) for training can be reported by UE.*  *Proposal 3: For NW-side model training, UE reports M2 L1-RSRPs measurements of fixed Set B (via Opt.2) to NW.*  *Proposal 4: For NW-side model training, UE reports Top-1 L1-RSRP and its associated beam (beam pair) index to NW (via Opt.1).*  *Proposal 5: For BM-Case2, study whether/how the time-domain information (e.g. timestamps) of collected data should be reported to NW.* |
| Samsung[24] | *Proposal 2. For BM-Case1 with a network-side AI/ML model, for the content of data collection, at least the following are identified:*  * Data (corresponding to model input)*  * L1-RSRP for at least all of Set B beams*  * Label*  * Beam ID for Set A (e.g., Top-1 beam ID)*  * L1-RSRP for at least all of Set A beams*  * FFS: assistance information (e.g., timestamp, UE speed, SNR, etc.)*  *Proposal 3. For BM-Case1 with a network-side AI/ML model, for the report container of data collection, the following observation is made:*  * L1 signaling*  * Data can be directly used for gNB-side model*  * Facilitates performance monitoring*  * RRC*  * Higher robustness in comparison with L1 signaling*  * Data can be directly used for gNB-side model*  * Requires more data storage in comparison with L1 signaling*  * User plane signaling*  * Higher robustness in comparison with L1 signaling*  * Data cannot be directly used for gNB-side model*  * Requires more data storage in comparison with L1 signaling*  *Proposal 4. For BM-Case1 with a network-side AI/ML model, for data collection, study the potential impact of the following aspect:*  * The handling/buffering for collected data before data collection reporting*  *Proposal 5. For BM-Case1 with a network-side AI/ML model, for the L1 signaling for the content of data collection, the following observation is made (for the options agreed in RAN1#112bis-e):*  * Option 1 requires additional support of L1 beam report with the measurement results of more than 4 beams in one reporting instance. Also, Option 1 may require additional support of L1 beam report with additional beam indication.*  * Option 2 requires additional support of L1 beam report with the measurement results of more than 4 beams in one reporting instance.*  * Option 3 can be supported with L1 beam report with beam indices only (without the corresponding L1-RSRP). This has addition specification impact. Alternatively, Option 3 can be achieved by existing beam reporting mechanism with the overhead of L1-RSRP.*  *Proposal 21. For BM-Case2 with a network-side AI/ML model, for the content of data collection, at least the following are identified on the top of BM-Case1:*  * Data (corresponding to model input)*  * Timestamp*  * Label*  * Timestamp*  *Proposal 22. For BM-Case2 with a network-side AI/ML model, for the report container of data collection, the observation is the same as BM-Case1:*   * *L1 signaling*   + *Data can be directly used for gNB-side model*   + *Facilitates performance monitoring* * *RRC*   + *Higher robustness in comparison with L1 signaling*   + *Data can be directly used for gNB-side model*   + *Requires more data storage in comparison with L1 signaling* * *User plane signaling*   + *Higher robustness in comparison with L1 signaling*   + *Data cannot be directly used for gNB-side model*   + *Requires more data storage in comparison with L1 signaling*   *Proposal 23. For BM-Case2 with a network-side AI/ML model, for the L1 signaling for the content of data collection, the following observation is made (for the options agreed in RAN1#112bis-e) on the top of BM-Case1:*  * For Option 1, Option 2 and Option 3, existing L1 beam report with timeRestrictionForChannelMeasurements set to 'configured' can be used to collect data with timestamp. This imposes restriction that timestamp of a RS measurement can only be derived based on the latest valid RS measurement occasion before the beam report instance.*  * Alternatively, for Option 1, Option 2 and Option 3, L1 beam report with the measurement results and flexible association of the corresponding timestamp may provide flexibility on time domain data collection, but may require additional specification efforts.* |
| Lenovo[27] | *Proposal 7: Study data collection procedure to support both UE-side and NW-side AI/ML model training and model update (if supported).*  * For UE-side model training, study to support UE triggered or UE initiated data collection procedure,*  * For NW-side model training, support to report larger number of beams in one beam report by considering beam report via MAC CE or RRC.* |
| CAICT[28] | *Proposal 1: NW should inform UE the contents of collected data and related time window for data collection for NW-side AI/ML model.*  *Proposal 2: Option1/2/3 could be supported as a framework for BM-Case 1 and BM-Case 2. Both L1 and RRC signalling should be used for related data transmission.* |
| KT[29] | *Proposal 1. For the model training at NW side, consider one shot and aperiodic reporting mechanism using both L1 and RRC signaling.* |
| MTK[31] | *Proposal 2: Study potential specification impact of using different Rx beam assumptions for AI/ML model inference for DL beam prediction.*  *Proposal 3: UE can determine its own Rx beam assumption used for measurement, regardless NW-side model’s Rx beam assumption.*  *Proposal 4: NW-side model’s Rx beam assumption can be implicitly indicated to UE through the corresponding reporting and RS configurations for data collection.*  *Proposal 8: For Opt.1 of the content of collected data for NW-side model, consider the following cases between the reported RSRP and indication of beams*  *• The beams of the reported M1 L1-RSRPs are the same as the beams of the reported indication*  *• The beams of the reported M1 L1-RSRPs are not the same as the beams of the reported indication*  *o M1 and the number of the indication of beams can be different*  *Proposal 9: For Opt.2 of the content of collected data for NW-side model, study the method and spec impact to quantize the reported M2 RSRP values.*  *Proposal 10: For the content of collected data for model inference of NW-side model, consider NW configures only Set B of beams for measurement and UE doesn’t report the indication for all the Set B of beams.*  *Proposal 11: For the content of collected data for model training of NW-side model, study mechanisms and spec impact of UE reports only the RSRP measurement of Set B of beams and label, without reporting the resource indication of Set B and/or Set A of beams.*  *Proposal 12: For the content of collected data for model inference of NW-side model, if data filtering/omitting is used, UE should use Opt.1 to report the measurement and the corresponding indication of the reported beams.*  *Proposal 13: For the content of collected data NW-side model, separately define the content for reporting the collected data for different data collection purposes and different SetA/SetB scenarios.*  *Proposal 14: For the content of collected data for NW-side model, study benefit and spec impact of UE determining M1,M2,M3 values for each reporting instance.*  *Proposal 15: For the content of collected data for NW-side model, further study defining the value of M1, M2, and M3 based on the following factors,*  *• Set A and Set B sizes*  *• data collection purposes*  *• model output design*  *• overhead reduction strategies*  *Proposal 16: For the data collection reporting mechanism for NW-side model, study how NW aligns with UE the following conditions for UE to report the corresponding content of data,*  *• Set A and Set B sizes*  *• data collection purposes*  *• model output design (e.g., label sizes, beam ID/RSRP)*  *• overhead reduction strategies (e.g., w/wo data omitting)*  *Proposal 18: For the data collection for NW-side model inference/training/monitoring for BM-Case2, study the benefit(s) and potential spec impact from the following two options of UE reporting as a starting point,*  *• Option1: NW configures separate report for each time instance in the observation window*  *• Option2: NW configures one report for all the time instances in the observation window*  *Proposal 18: For the data collection for NW-side model inference/training/monitoring for BM-Case2, study the benefit(s) and potential spec impact from the following two options of UE reporting as a starting point,*  *• Option1: NW configures separate report for each time instance in the observation window*  *• Option2: NW configures one report for all the time instances in the observation window*  *Proposal 19: For data collection for NW-side AI/ML model, study spec impact of reporting overhead reduction with quantizing L1-RSRP and/or normalized L1-RSRP measurement with lower number of bits than the current spec.* |
|  |  |

###### **Mod’s assessment**

There are various aspects discussed in the tdocs, e.g.,

* The range of M1/M2/M3 for data collection
* Signaling/configuration of beams/RS
* Down-selection / combination of Option1/2/3
* Overhead reduction
* Signaling container, e.g., L1 signaling, RRC, user plane data
* Common understanding the Rx beam used for data collection
* Reporting of information related to Rx beam
* Data buffering/Measurement window
* RS enhancement for better measurement accuracy
* Advance receiver for beam measurement
* …

**Value of M1/M2/M3**

At least 4 companies (e.g., H3C, IDC, Ericsson, MTK) have some discussions related to the range of M1/M2/M3 in Option1/2/3 for data collection.

* **Mod’s assessment:** We have agreed that M1/M2/M3 can be larger than 4, which is a key information different from the current spec (where only up to 4 beams can be reported). Regarding the exact range of M1/M2/M3, if the group can converge on some value, then we can make some agreement/conclusion in R18 study item. Otherwise, it can be discussed later or in R19 WI (if any).
* **Mod’s suggestion:** As there are only 4 companies (or more?) discussing this issue in tdocs, moderator suggests not to discuss the issue in this meeting.
* **Note:** If the LS discussion in AI 9.2.1 needs some information about the values of M1/M2/M3 (e.g., due to the discussion of data size), the group can discuss it in AI 9.2.1
* **Related proposals in tdocs**
  + H3C: Proposal 5
  + IDC: Proposal 30
  + Ericsson: Proposal 6
  + MTK: Proposal 15

**Configuration/Signaling for data collection**

At least 8 companies (e.g., H3C, Huawei, vivo, ZTE, Fujitsu, Xiaomi, CAICT, Samsung) have discussions related to the configuration/signaling of data collection (e.g., RS/beam/resource configuration, reporting configuration, measurement window, …)

* **Mod’s assessment:** Although there are many companies discussing this issue, the proposals are quite divergent. It seems unlikely to have some stable proposal(s) for these details in the current stage. Moreover, the detailed configuration/signaling can be discussed in WI.
* **Mod’s suggestion:** No discussion on the detailed configuration/signaling in this meeting.
* **Related proposals in tdocs**
  + H3C: Proposal 6
  + Huawei: Proposal 9
  + Vivo: Proposal 18
  + ZTE: Proposal 3
  + Fujitsu: Proposal 2
  + Xiaomi: Proposal 16
  + CAICT: Proposal 1
  + Samsung: Proposal 23

**Down-selection/prioritization/combination of Option1/2/3**

At least 7 companies (e.g., Huawei, Spreadtrum, IDC, Intel, Nokia, OPPO, CAICT) have some discussion related to the down-selection/prioritization/combination of Option1/2/3. There are also some companies (e.g., Samsung, Apple, Xiaomi) discussing contents of the collected data.

* **Mod’s assessment:** Each option has its own proponents and associated use cases. Meanwhile, down-selection/prioritization/combination of Option1/2/3 is related to contents of the collected data. As the LS discussion will be in AI 9.2.1, this agenda item should avoid the duplicated discussion.
* **Mod’s suggestion:** Not discuss the contents of the collected data (including the down-selection/prioritization/combination of Option1/2/3) in AI 9.2.3.2 this meeting.
* **Related proposals in tdocs**
  + Huawei: Proposal 10
  + Spreadtrum: Proposal 4
  + IDC: Proposal 29
  + Intel: Proposal 9
  + Nokia: Proposal 22
  + OPPO: Proposal 3, 4
  + CAICT: Proposal 2
  + Samsung: Proposal 2, 21
  + Apple: Proposal 6
  + Xiaomi: Proposal 22

**Signaling container of reporting**

At least there are 6 companies (e.g., Huawei, CMCC, CAICT, Samsung, KT, Lenovo) have some discussions related to the signaling container of data collection for training/monitoring (e.g., L1 signaling, RRC, MAC CE, User plane data).

* **Mod’s assessment:** Whether and how L1 signaling is supported for data collection or not is within RAN1 scope, and it is better to have a clear conclusion/agreement in study item.
* **Mod’s suggestion:** Let’s focus on whether L1 signaling is supported for data collection for training/monitoring. How higher layer signaling based solution is supported is up to other WGs. Thus, a proposal is suggested for discussion.
* **Related proposals in tdocs**
  + Huawei: Proposal 11
  + CMCC: Proposal 4
  + CAICT: Proposal 2
  + KT: Proposal 1
  + Lenovo: Proposal 7
  + Samsung: Proposal 3, 22

**Overhead reduction**

At least 15 companies (e.g., Futurewei, vivo, ZTE, CATT, NEC, CMCC, DCM, Nokia, Ericsson, MTK, Apple, OPPO, Samsung, ETRI, Lenovo) continue to discuss different methods to reduce the overhead (or related to the overhead reduction).

* **Mod’s assessment:** A proposal was suggested and discussed for overhead reduction in the last meeting. Although most companies supported the proposal, no consensus was made since two companies commented that it was not needed to discuss this issue. Since there are 15 supporting companies for overhead reduction, it should be an important issue from the views of most companies. Thus, it is worth of another round(s) of discussions.
* **Mod’s suggestion:** As there are still many companies suggesting method(s) for overhead reduction, moderator suggested to continue discussion. A revision based on the proposal of the last meeting is suggested for discussion.
* **Related proposals/observation on data collection for general purpose**
  + Futurewei: Proposal 1
  + Vivo: Proposal 21
  + ZTE: Proposal 5,7
  + CATT: Proposal 6
  + NEC: Proposal 5
  + CMCC: Proposal 5
  + DCM: Observation 5
  + Nokia: Proposal 23
  + Ericsson: Proposal 7
  + MTK: Proposal 19
  + Apple: Proposal 4
* **Related proposals/observation dedicated to model inference** 
  + Vivo: Proposal 31
  + ZTE: Proposal 10
  + Ericsson: Proposal 10
  + CMCC: Proposal 16
  + DCM: Proposal 9
  + OPPO: Proposal 12
  + Samsung: Proposal 7
  + ETRI: Proposal 26
  + Lenovo: Proposal 2
  + MTK: Proposal 23

**Rx assumption for DL Tx beam prediction**

At least 5 companies (e.g., Futurewei, Spreadtrum, CMCC, Nokia, MTK) continue discussing how to ensure the same Rx assumption between NW and UE for DL Tx beam prediction at NW side.

* **Mod’s assessment:** In the last meeting, a proposal was suggested and discussed. There were 8 supporting companies (CMCC, Spreadtrum, IDC, H3C, Futurewei, ZTE, DCM, Xiaomi) whereas 8 companies (CATT, LG, Lenovo, Ericsson, QC, Google, Samsung, Huawei) didn’t support it. Based on the tdocs submitted to this meeting, there are 2 more supporting companies (Nokia, MTK). If this issue is valid (no consensus so far), the common Rx assumption should be a critical issue for AI/ML based DL T x beam prediction. As 18 companies are involved in this discussion (10 vs 8). it is worthy of having another chance for further discussion.
* **Mod’s suggestion:** A proposal based on the one of the last meeting is suggested for further discussion. One of the following outputs is expected this meeting
  + Agree to study some mechanism to ensure the common Rx beam assumption
  + Conclude that there is no issue (i.e., legacy TCI framework can ensure the common Rx beam assumption)
* **Note:** There are also some proposals for the signaling/assistance information to ensure the common Rx assumption (e.g., DCM Proposal 8). The detailed design can be discussed after the group achieve consensus on this issue.
* **Related proposals in tdocs**
  + Futurewei: Proposal 2
  + Spreadtrum: Proposal 6
  + CMCC: Proposal 3
  + Nokia: Proposal 24
  + MTK: Proposal 2

**Rx beam reporting for DL beam pair prediction**

Some companies are discussing the reporting of Rx beam for DL beam pair prediction.

* **Mod’s assessment:** It is quite controversial on whether to support DL beam pair prediction or not. It is better to discuss the detailed mechanism dedicated to DL beam pair prediction after the group have consensus on the support of beam pair prediction.
* **Mod’s suggestion:** Let’s first discuss and make agreement/conclusion on whether to support beam pair prediction. After that, we can come back to this issue if needed.

**Other issues**

There are also some proposals for various other aspects

* **Mod’s assessment:** Most of these proposals are only suggested by 1 or 2 companies. Thus, from moderator’s perspective, it is difficult to formulate some stable proposal(s).
* **Mod’s suggestion:** No discussion until there are more supporting companies.

###### Proposal 2.2.1 Closed

Data collection for training

* L1
  + Huawei: Proposal 11
  + CMCC: Proposal 4
  + CAICT: Proposal 2
  + KT: Proposal 1
* RRC
  + Huawei: Proposal 11
  + Lenovo: Proposal 7
  + CAICT: Proposal 2
  + KT: Proposal 1
* MAC
  + Lenovo: Proposal 7

***Proposal 2.2.1: Regarding data collection for NW-side AI/ML model of BM-Case1 and BM-Case2, the following reporting signaling for beam-specific aspects are considered:***

* ***L1 signaling to report the collected data***
* ***Higher-layer signaling to report the collected data*** 
  + ***At least not applicable to AI/ML model inference***
* ***Note1: higher layer signaling design is up to other WG(s)***
* ***Note2: Whether each signaling applicable to each ~~LMC~~ LCM purpose is a separate discussion***
* ***Note3: The legacy signaling principle (e.g. RSRP reporting for L1) can be re-used***

|  |  |
| --- | --- |
| Company | Comments |
| Mod | The group agreed to study the spec impact of L1 reporting signaling for AI model inference in RAN1#111. |
| LG | RAN1 should focus on L1 signaling based solution. In this regard, we don’t see any delta compared to previous agreement from this proposal. |
| Xiaomi | For performance monitoring, there was an agreement copied below. It is better to remove the repeated part.  While for model training, we are fine with it and suggest to add a note to the 1st sub-bullet: L1 signaling can only report a part of sample(s). e.g., only report the label including Top-K beam ID.  ***Agreement***  ***Regarding NW-side model monitoring for a network-side AI/ML model of BM-Case1 and BM-Case2, study the necessity and the potential specification impacts from the following aspects:***   * ***UE reporting of beam measurement(s) based on a set of beams indicated by gNB*** * ***Signaling, e.g., RRC-based, L1-based*** * ***Note: Performance and UE complexity, power consumption should be considered*** |
| NTT DOCOMO | Share the same view with LG. |
| CATT | We support the proposal.  Regarding data for training, we think both L1 signaling and higher-layer signaling can be considered. L1 signaling can be used to collect a small number of samples and higher-layer signaling can be used to collect a large number of samples  Regarding data for performance monitoring, at least L1 signaling can be supported, whether the higher-layer signaling can be considered will depend on the specific performance calculation method (e.g., a number of monitoring instances to ensure the robustness) and latency requirement. |
| ZTE | Regarding the reporting of collected data, we prefer L1 signaling to be used for model monitoring and high-layer signaling to be used for model training due to their different latency requirements. |
| Samsung | The aspect of model inference is missing from this proposal. Due to the stringent latency requirement, L1 signaling is preferred for the report of collected data for model inference.  For the formulation of the proposal, it is also suggested to make the statement as an outcome of study.  Hence, we suggest the following change to the FL proposal.  ***Proposal 2.2.1: Regarding data collection for NW-side AI/ML model of BM-Case1 and BM-Case2, ~~study~~ consider the potential specification enhancement to support the following:***   * ***L1 signaling to report the collected data for training*** * ***Higher-layer signaling to report the collected data for training*** * ***L1 signaling to report the collected data for model inference*** * ***L1 signaling to report the collected data for performance monitoring*** * ***Higher-layer signaling to report the collected data for performance monitoring*** * ***Note: higher layer signaling design is up to other WG(s)*** |
| Hw/HiSi | We agree with that the **signaling** for both mechanisms should be studied.  Regarding the bullets for higher layer signaling, can the intention please be clarified? On one hand it is said that the RAN1 should study the spec impact but on the other hand there is a note saying that the design is up other WGs.  Since we already have an agreement for inference, we think it would be good to state in the main bullet that this is for training and monitoring.  Rather than the spec impact, we think it should be specified that the signaling should be studied. In our view, the spec impact for the reporting mechanism is not a given and then legacy reporting could be re-used and if there is additional spec impact could be started at a later stage.  Based on above comments we suggest to make the following modifications  ***Proposal 2.2.1: Regarding data collection for training/monitoring for NW-side AI/ML model of BM-Case1 and BM-Case2, study the signaling ~~specification impact~~ (to support the following:***   * ***L1 signaling to report the collected data ~~for training~~*** * ***Higher-layer signaling to report the collected ~~data for training~~*** * ***~~L1 signaling to report the collected data for performance monitoring~~*** * ***~~Higher-layer signaling to report the collected data for performance monitoring~~***   ***Note: ~~higher layer signaling design is up to other WG(s)~~***  ***Note: the legacy reporting mechanism can be re-used*** |
| QC | Agree with LG. |
| Lenovo | We are fine with this proposal. At least for data collection for model train at the NW side, higher layers can be considered. |
| NEC | In our view, at least L1 signaling should be supported for model inference, model monitoring and model training. L3 signaling can be considered for model training due to specific latency requirement. |
| Futurewei | As we discussed earlier (Section 7.2.2 of this document), RAN2 made agreement in RAN2#121bis to deprioritize online training. Meanwhile, based on the discussion in previous meetings, most companies suggest prioritizing offline training. In this case, we may not need to use L1 signaling to report the collected data for training; it is too signaling-heavy. |
| Ericsson | We agree with Samsung that since this is the final meeting, we should avoid proposals on “what to study”. Our view is a combination of Huawei and Samsung proposal.  ***Proposal 2.2.1: Regarding data collection for training/monitoring for NW-side AI/ML model of BM-Case1 and BM-Case2, consider the potential specification enhancement to support the following:***   * ***L1 signaling to report the collected data ~~for training~~*** * ***Higher-layer (e.g. RRC) signaling to report the collected ~~data for training~~*** * ***~~L1 signaling to report the collected data for performance monitoring~~*** * ***~~Higher-layer signaling to report the collected data for performance monitoring~~***   ***Note: ~~higher layer signaling design is up to other WG(s)~~***  ***Note: the legacy reporting mechanism can be re-used*** |
| Nokia/NSB | In general, ok with the updated version shared by E///.  Suggest to delete “training/monitoring” as RAN1 data collection discussions can be generic as we did in last few meetings. |
| Spreadtrum | We are fine with this proposal. We think L1 signaling can be used for model monitoring/inference. |
| Mod | The proposal is updated based on the comments.  Note: I didn’t make reply to each companies’ comment in the summary. I will try to find the delegates and have some F2F talk, which may be more efficient (This note is applicable to all proposals) |
| MediaTek | We are fine with the current proposal. |
| CMCC | Fine with this proposal. |
| Fujitsu | Fine with this proposal. |
| Google | We think it is better to clarify whether the L1 signaling is a control signaling from the NW to configure the report, or to trigger the report or a report from the UE or all above. If this is based on the beam report framework, the key problem is what additional control information is needed from the network, instead of whether the signaling is a L1 or higher layer signaling. |
| Xiaomi | ‘LMC’ in the note should be ‘LCM’ |
| NEC | Support the updated proposal. |
| ZTE | The last note seems contradict with the first and second bullets. In the first and second bullets, it is said that both L1 signaling and higher-layer signaling are supported or considered to be supported for the reporting of the collected data. But in the last note, it is said whether L1 signaling or higher-layer signaling is applicable/supported is a separate discussion. Thus, it is quite confusing for the intention of this proposal. |
| HW/HiSi | Almost ok.  We have not seen yet any indication to deviate from the legacy reporting principles. To keep the future discussion focused, it would be good to capture this is a note.  ***Proposal 2.2.1: Regarding data collection for NW-side AI/ML model of BM-Case1 and BM-Case2, the following signaling for beam-specific aspects are considered:***   * ***L1 signaling to report the collected data ~~for training~~*** * ***Higher-layer signaling to report the collected data ~~for training~~***   + ***At least not applicable to AI/ML model inference*** * ***~~L1 signaling to report the collected data for performance monitoring~~*** * ***~~Higher-layer signaling to report the collected data for performance monitoring~~*** * ***Note: higher layer signaling design is up to other WG(s)*** * ***Note: Whether each signaling applicable to each LMC purpose is a separate discussion***   ***Note: The legacy signaling principle (e.g. RSRP reporting for L1) can be re-used*** |
| Mod | The proposal is updated based on the comments from Google/Xiaomi/Huawei  @ZTE: The note is talking about the applicability to each LMC purpose. Thus, it seems no confliction. |
| ZTE | It’s not clear the specific meaning of ‘signaling principle’. If the last note is indeed needed, we suggest the following revision.   * ***Note3: The legacy signaling ~~principle~~ (e.g. RSRP reporting for L1) ~~can~~ may be re-used*** |
| Mod | Closed. Please see the corresponding agreement made in online session. |

###### Proposal 2.2.2 Closed

* Omit some data (be selective on data)
  + Futurewei: Proposal 1
  + Vivo: Proposal 21
  + ZTE: Proposal 7
  + ZTE: Proposal 5 (implicit reporting of beam ID)
  + CATT: Proposal 6 (including implicit reporting of beam ID)
  + NEC: Proposal 5
  + CMCC: Proposal 5
  + DCM: Observation 5
  + Nokia: Proposal 23
* Compression (including larger quantization step, differential quantization, exploiting correlation between samples)
  + Futurewei: Proposal 1
  + ZTE: Proposal 7
  + Ericsson: Proposal 7
  + CATT: Proposal 6
  + CMCC: Proposal 5
  + MTK: Proposal 19
  + Apple: Proposal 4

***Proposal 2.2.2: Regarding data collection for NW-side AI/ML model of BM-Case1 and BM-Case2, the following approaches have been studied by companies for overhead reduction***

* ***the omission/selection of collected data***
* ***the compression of collected data***
* ***Note1: For the different purposes of data collection, the overhead reduction mechanisms and corresponding specification impacts may be different.***
* ***Note2: Support of any mechanism(s) (if necessary) for each LCM purpose and the potential spec impact (if any) are separate discussions***
* ***Note 3: UE complexity and power consumption should be considered.***

|  |  |
| --- | --- |
| Company | Comments |
| Mod | To Futurewei: As most tdocs are using “omit”, the proposal uses “omit” rather than “be selective on”. |
| LG | Fine |
| Xiaomi | Support |
| NTT DOCOMO | Compression includes omission in our understanding. To avoid the overlap between first bullet and second bullet, we prefer to update the second bullet as follows.  ***Proposal 2.2.2: Regarding data collection for NW-side AI/ML model of BM-Case1 and BM-Case2, study specification impacts for overhead reduction by considering the following approaches:***   * ***Whether/how to omit the field of some data*** * ***Whether/how to ~~compress the~~ reduce bitwidth of the field corresponding to some reported content*** * ***Note: For the different purposes of data collection, the overhead reduction mechanisms and corresponding specification impacts may be different.*** |
| CATT | The data should be clarified further. In our view, the data at least refers to the L1-RSRP and beam index. One approach is to omit reporting L1-RSRP of some beams. Another approach is to omit reporting beam index of some beams. |
| ZTE | Support. Both data omission (e.g., according to data quality) and compression on reported content (e.g., lager quantization step size, threshold-based measurement report, measurement report with a pre-defined order) can be studied for overhead reduction. |
| Samsung | The intention of this proposal is understandable. However, the current formulation of the proposal includes ‘study’, ‘whether/how’, which is very difficult to reflect the study outcome of overhead reduction.  In order to make the proposal more meaningful, we suggest the following change to the FL proposal.  ***Proposal 2.2.2: Regarding data collection for NW-side AI/ML model of BM-Case1 and BM-Case2, ~~study~~ consider specification ~~impacts~~ enhancement for overhead reduction by ~~considering~~ the following approaches:***   * ***~~Whether/how to omit some data~~ the omission of collected data*** * ***~~Whether/how to compress the reported content~~ the compression of collected data*** * ***Note: For the different purposes of data collection, the overhead reduction mechanisms and corresponding specification impacts may be different.*** |
| HW/HiSi | Not support.  The need and justification for overhead reduction has not really been given for yet. What we have seen from submitted papers there are many diversified schemes and it is also not either explained whether the overhead reduction is intended for training and inference.  Given that the current specification already allows to report a subset of the measurement beams, e.g. the strongest beams and differential RSRPs, we fail to see what other schemes could bring. Proponents should clarify more on this. |
| QC | This is an incremental proposal on top of what has been already agreed:  **Agreement**  ***Regarding data collection for NW-side AI/ML model, study necessity, benefits and beam-management-specific potential specification impact from RAN1 point of view on the following additional aspects***   * ***Mechanism related to the reporting*** * ***Additional information for content of the reporting***   + ***FFS: Information associated with or configured for the reported data samples, e.g., timestamps, SNR, data quality, etc.*** * ***Reporting overhead reduction*** * ***Note1: non-3GPP based solution is a separate issue.*** * ***Note2: The framework corresponding to higher layer(s) are up to the associated WG(s)*** * ***Note 3: Overhead, UE complexity and power consumption should be considered***   Hence, we do not see the need for this proposal. As a compromise, we suggest the following changes:  ***Updated Proposal 2.2.2: Regarding data collection for NW-side AI/ML model of BM-Case1 and BM-Case2, study ~~specification impacts~~ necessity, benefits and beam-management-specific potential specification impact from RAN1 point of view (if any) for overhead reduction by considering the following approaches:***   * ***Whether/how to omit some data*** * ***Whether/how to compress the reported content*** * ***Note 1: For the different purposes of data collection, the overhead reduction mechanisms and corresponding specification impacts may be different.*** * ***Note 2: UE complexity and power consumption should be considered.*** |
| Lenovo | We suggest providing some examples to shown how to ‘omit’ or ‘compress’ the data. |
| NEC | Fine |
| Ericsson | We should avoid what to study at this stage. Support the update from Samsung.  We think the first bullet can be useful to avoid reporting duplicated samples (e.g. samples that are collected within one channel coherence time interval) from the UE, to save reporting overhead. ´ |
| Nokia/NSB | The updated version from Samsung looks Ok in general. |
| Spreadtrum | We are generally fine with this proposal. |
| Sony | Support |
| Mod | The proposal is updated |
| MediaTek | We support this proposal. |
| CMCC | Generally fine with this proposal. Overhead of training, inference, monitoring is different, where overhead of training can be large if RSRP of beams in setA is reported. Overhead of inference can be small if number of beams in set B is small. Overhead of monitoring depend on performance KPI. It is not clear which phase is the overhead reduction applied to. |
| Fujitsu | Fine to the updated proposal. |
| Google | We think overhead reduction for data collection should be studied after we determine what needs to be reported. |
| Xiaomi | Fine |
| NEC | OK with the updated proposal. |
| Mod | @CMCC: Note2 is to say, we will discuss the detailed mechanisms (if needed) for each LCM purpose separately, |
| Mod | Closed. Please see the corresponding agreement in Chair’s note |

###### Proposal 2.2.3 Closed

The related proposals in tdocs are as below:

* Futurewei: Proposal 2
* Spreadtrum: Proposal 6
* CMCC: Proposal 3
* Nokia: Proposal 24
* MTK: Proposal 2

***Proposal 2.2.3: Regarding data collection for NW-side AI/ML model of BM-Case1 and BM-Case2, study potential specification impact on the following aspect for DL Tx beam prediction***

* ***Mechanism to ensure ~~the common understanding between NW and UE on~~ Rx beam assumption for reported measurement result(s) to be the same for both AI/ML model training and model/ML inference***

Check whether there is any issue on the common understanding between NW and UE for the following Rx assumption for measurement:

* best Rx beam
* fixed Rx beam
* random Rx beam

|  |  |
| --- | --- |
|  | Companies |
| Is an issue | New H3C (for Fixed RX beam) |
| Not an issue |  |

|  |  |
| --- | --- |
| Company | Comments |
| Mod | We need to make some conclusion/agreement on this issue. One of the following outputs is expected this meeting   * Agree to study some mechanism to ensure the common Rx beam assumption * Conclude that there is no issue (i.e., legacy TCI framework can ensure the common Rx beam assumption) |
| LG | Unclear on what “common understanding on UE Rx beam assumption” refers to. |
| Xiaomi | Model with different Rx beam assumption may have different performance, thus it is better to have common understanding on Rx beam assumption between UE and gNB. |
| NTT DOCOMO | Support the proposal. There are diverse Rx beam determination mechanisms in 9.2.3.1. For example, best Rx beam for each Tx beam or best Rx beam for the best Set B/C Tx beam. Some companies claim that QCL can be used for common understanding. However, it does not guarantee the measurements in one reporting instance are measured with the same Rx beam. For instance, if Rx beam sweeping is performed with repetition of CSI-RS, the reported measurements could be same or different for beam measurements corresponding to different Tx beams. To avoid the consistency over training and inference for NW side model, we think Rx beam assumption should be aligned between NW and UE. |
| CATT | We are ok with this proposal. It may need some enhancement mechanism to ensure the common understanding between NW and UE on RX beam assumption at lease for DL Tx beam prediction with a specific Rx beam assumption. |
| ZTE | Legacy TCI framework can ensure the common Rx beam assumption. In the current specification, the spatial Rx parameter of the UE is indicated by the higher layer parameter of quasi co-location relationship with respect to QCL type set to type D. That is, which Rx beam to use for beam measurement can be indicated by the QCL configuration from the NW side. Therefore, the common understanding between NW and UE on Rx beam assumption for reported measurement results can be ensured by QCL indication/configuration in the legacy TCI framework. For example, assuming a Rx beam assumption that only one Rx beam is used for the measurement of different Tx beams in Set A/B, all beams in the measured beam set can be configured with the same QCL type D relationship, which implies that the same Rx beam should be used for beam measurement. |
| Samsung | In our understanding, this proposal is associated with the beam indication of Set A beam but not in the Set B. It is suggested to combined the discussion of these two aspects. |
| Hw/HiSi | Not support.  The Rx beam is up to UE implementation and there is no need to disclose this information to gNB, neither in an implicit or direct method. The evaluations in 9.2.3.1 have shown that such a common understanding is not needed for the Tx beam prediction. For beam pair prediction, on the other hand, it would be different. |
| QC | Not support. Unclear what “common understanding between NW and UE on Rx beam assumption” means. Additionally, there are quite a few simulations provided by multiple companies showing that for DL Tx beam prediction, the generalization performance across different UE codebooks is very good, hence no need for NW to have such information. |
| Lenovo | Can the “Rx beam assumption” be a specific Rx beam or among all the Rx beams? |
| Ericsson | Support the intention of the proposal. The evaluations captured below have shown that a Random UE RX beam selection can impact the TX-beam prediction heavily. Proposed update below:   * ***Mechanism to indicate ~~ensure the common understanding between NW and~~ UE ~~on~~ Rx beam assumption for reported measurement result(s)*** * ***Indication should not disclose proprietary information (e.g. UE can indicate that it has not used a random RX beam)***   Observation  At least for BM-Case1 when Set B is a subset of Set A, and for DL Tx beam prediction, with the measurements of the “best” Rx beam with exhaustive beam sweeping for each model input sample, AI/ML provides the better performance than with measurements of random Rx beam(s).   * Evaluation results from [8 sources: vivo, Nokia, Fujitsu, Samsung Lenovo, Huawei/HiSi, Ericsson, MediaTek] show [25%~50%] degradation with random Rx beam(s) comparing with the “best” Rx beam in terms of Top-1 prediction accuracy. * Evaluation results from [1 source: CATT] show **about 6%** degradation with measurement of random Rx compared with measurement of best Rx in term of Top-1 beam prediction accuracy.   Comparing performance with non-AI baseline option 2 (based on the measurement from Set B of beams), with measurements of random Rx beam(s) as AI/ML inputs:   * Evaluation results from [5 sources: MediaTek, Fujitsu, vivo, Nokia Samsung] show that AI/ML can still provide [7%~44%] beam prediction accuracy gain in terms of Top-1 beam prediction accuracy.   Note: In both training and inference, measurements of random Rx beams are used as AI/ML inputs. |
| Nokia/NSB | Agree with the E/// comment. |
| Spreadtrum | Not support. First, the Rx beam should belong to the UE implementation, so gNB should not get Rx information; Secondly, according to the simulation results in 9.2.3.1, better performance can be obtained by using the optimal (quasi-optimal) Rx beam, which can be used by UE for receiving. There is therefore no need to introduce new mechanisms。 |
| Mod | The proposal is updated. Let’s wait to hear more views |
| MediaTek | We are open to study this for the following reasons:  (1) This will not affect UE implementation, like ZTE mentioned, the common understanding of Rx beam assumption can be achieved by NW configuring QCL Type D for beams in Set B in the legacy TCI framework. Our understanding is that this proposal is for NW to indicate a **mechanism** for UE to determine which Rx beam to use for measuring Set B of beams. This mechanism can be Best-Rx (exhaustive search), random Rx (if such model with good performance can be found) or different choices of quasi-Rx beams. How to form a random Rx beam or quasi-Rx beam codebook is still up to UE’s implementation. For example, such signaling cannot be “NW indicating UE to use Rx beam #N and/or to use a certain Rx beam direction”. Therefore, we support what E/// proposed to include the following note in the proposal:  • Indication should not disclose proprietary information (e.g. UE can indicate that it has not used a random RX beam)  (2) The problem of using legacy TCI framework for aligning Rx beam assumption is that it might induce large RRC signaling overhead. Assuming a NW-side model is using quasi-Rx beam assumption and requires UE measuring Set B by using the best Rx of “the previously used Tx beam”. Since “the previously used Tx beam” is changing overtime, for each round of beam management procedure, NW needs to send an additional RRC signal to configure that the TCI state of each beam in Set B is TypeD QCLed with “the previously used Tx beam”. Such RRC signaling mechanism is unnecessarily complex and should be avoid, considering we just want a simple alignment of Rx beam assumption. Thus, we believe it is worth to study if we can have a more efficient way to align such Rx beam assumption between UE and NW. |
| CMCC | Support. Different Rx beam assumption corresponds to different repetition number of Tx beams and different Rx beam selection. For example, if NW side model is trained with best Rx beam, and each Tx beam is repeated 4 times during inference, UE needs to report RSRP of best Rx beam instead of random Rx beam. If NW side model is trained with quasi-optimal Rx beam, and each Tx beam is transmitted one time during inference, UE needs to report RSRP of quasi-optimal Rx beam instead of random Rx beam.When NW side model is trained with one Rx beam assumption, but inference with another Rx beam assumption, the performance may not be good. |
| Fujitsu | The updated proposal seems to want to ensure the consistency on Rx beam assumption across the training and inference. But such consistency should be ensured not only on the Rx beam assumption but also on Tx beam assumption for UE side model. The intension of this proposal is not clear. |
| Google | Do not support. Rx beam should be up to UE implementation. Spec should not require the UE to disclose the Rx beam. |
| Xiaomi | It is better to add a note to explain ‘Rx beam assumption’. We think it means best Rx beam, fixed Rx beam or random Rx beam. |
| CATT | We need to clarify in which cases the Rx beam assumption should be the same in model training and model inference. For example, if best Rx beam assumption is used in model training, and quasi-optimal (with 80% or above best Rx beam) assumption is used in model inference, the performance degradation is very little. |
| ZTE | For the subbullet, whether the reported measurement results are used for model training or model inference is transparent to the UE, and thus we don't need to mention that point. We assume the intention of this proposal is to study the potential specification impact (if needed) on the indication/assurance of Rx beam assumption for reported measurement results, regardless of model training or model inference. |
| HW/HiSi | Not support.  The Rx beam is information is proprietary for the UE implementation and a risk to disclose this should be avoided. We agree with QC that simulations results indicate that the such information is not necessary. And if some common understanding should be achieved anyway, then, as ZTE commented in the first round, the legacy mechanism is well suited for this. |
| Mod | A table is added for proposal 2.2.3 to collect views on whether it is a valid issue. Based on the input, maybe we can conclude whether it is issue or not. |
| Apple | We cannot support this proposal. We don’t believe the random Rx beam evaluation is a good reason at all. In conventional BM approach, is random Rx beam selection used? If not, then why does one have to assume AI-BM use inferior solutions for Rx beam than the conventional approach? We don’t agree on working manufactured problems. |
| InterDigital | We do not support this proposal. 1. Beam pair prediction can achieve prediction with different Rx beams 2. Tx beam prediction is also possible if the model is generalized with multiple Rx beams. |
| CMCC | In our understanding, Rx beam assumption includes best Rx beam, quasi-optimal Rx beam, specific Rx beam, random Rx beam.  Checking evaluation results in 9.2.3.1, AI/ML-based model has good performance when the quasi-optimal Rx beam selection can select the genie best Rx beam with probability more than 75% or 80%. If Rx beam is up to UE implementation and Rx beam assumption considers only best Rx beam and quasi-optimal Rx beam, generalization of Rx beam seems fine, gNB does not need indicate Rx beam assumption. If Rx beam is up to UE implementation and Rx beam assumption also considers specific Rx beam and random Rx beam, for a UE which always use specific Rx beam, the prediction accuracy may not be good. |
| New H3C | for fixed RX beam, there is an issue and UE should keep the beam not changed between training and inference stage |
| Mod | Based on the inputs and offline discussion, this issue is still quite controversial. We have no time to have some detailed discussion on it in this meeting. |

## UE-side AI model

In previous RAN1 meeting(s), the agreement(s)/conclusion(s) were made as below:

|  |
| --- |
| **RAN1#111**  Agreement  Regarding the data collection for AI/ML model training at UE side, study the potential specification impact considering the following additional aspects.   * Whether and how to initiate data collection * Configurations, e.g., configuration related to set A and/or Set B, information on association/mapping of Set A and Set B * Assistance information from Network to UE (If supported) * Other aspect(s) is not precluded   **RAN1#112bis-e**  Agreement  Regarding the data collection at UE side for UE-side AI/ML model, study the potential specification impact of UE reporting to network from the following aspect   * Supported/preferred configurations of DL RS transmission * Other aspect(s) is not precluded   Agreement  Regarding the data collection at UE side for UE-side AI/ML model, study the potential specification impact (if any) to initiate/trigger data collection from RAN1 point of view by considering the following options as a starting point   * Option 1: data collection initiated/triggered by configuration from NW * Option 2: request from UE for data collection   + FFS: details |

The related proposals in tdocs are copied as below:

|  |  |
| --- | --- |
| H3C [1] | *Proposal7: Regarding data collection for UE-side AI/ML model training, trigger the data collection procedure based on UE request with high-layer signaling.* |
| Spreadtrum[5] | *Proposal 3：For UE-side model, support UE to request the data collection and report training-related information, such as expected measurement resources, etc.* |
| IDC[6] | *Proposal 26: Data collection initiation/trigger at the UE side for a UE-sided model should be based on some input from the UE.*  *Proposal 27: For UE request of data collection in a UE sided AI/ML model, support gNB confirmation procedure to avoid misunderstanding between UE and gNB.* |
| Vivo[7] | *Proposal 23: Regarding the data collection for AI/ML model training at UE side, study potential specification impact on resource configuration:*  *• Specific beam pair resource configuration for Set A*  *• Enhanced P3+P2 resource configuration that Rx beam assumption of P2 resource measurement is the best Rx beam searched from P3 procedure for performance improvement*  *Proposal 24: Regarding the data collection for AI/ML model training at UE side, study potential specification impact on assistance information:*  *• Proprietary processed Tx beam information as assistance information from NW to UE*  *Proposal 25: Regarding the data collection for AI/ML model training at UE side, study the potential specification impact on request signaling:*  *• Resource request signaling for data collection from UE to NW*  *- Beam pair resources request*  *- P3 and/or P2 beam sweeping resources request*  *• Minimum resource number request for data collection from UE to NW*  *- Minimum number of beams requested for model training w or w/o resource request signaling*  *- Minimum number of repetitions requested for model training w or w/o resource request signaling* |
| Intel[9] | *Proposal 2: For UE side AI/ML model, support UE triggered data collection where the triggering can be with respect to pre-configured set of TCI states and reference signal transmissions.* |
| Ericsson[11] | *Observation 4 The UE can initiate data collection based on the received beam ID* |
| Google[12] | *Proposal 18: For AI/ML in UE side, study the potential enhancement to maintain the same understanding between the gNB and UE with regard to the reported beam information based on a beam-codebook similar to CSI feedback based on a codebook*  *• The UE can report a beam matrix indicator (BMI) based on the beam-codebook* |
| LG[13] | *Observation #1: In current specification, UE cannot assume that Tx beam applied for a NZP CSI-RS resource will be unchanged in different transmission instances.*  *Proposal #2: For data collection at UE-side model, it should be studied on how to ensure the same Tx beam for different transmission instances of same or different CSI-RS resources.*  *Proposal #4: Consider UE assistance/reporting for determining Set A, e.g. UE to report preferred Set A among candidate beams of Set A.* |
| CATT[14] | *Proposal 2: For DL beam pair prediction with a UE-side model, study the following aspects:*  *• For model training at NW side, study how to report relative Rx beam information when preserving sensitive proprietary information*  *• For model training at UE side, study how to send/report relative Tx beam information when preserving sensitive proprietary information.*  *Proposal 3: For DL beam pair prediction with model training at NW side and inference at UE side, study how to align the mapping rule for Tx beam ID and Rx beam ID between the NW and UE.*  *Proposal 7: Regarding the training data collection for AI/ML model training at UE side, study the potential specification impact on the following aspects:*  *• RS transmission to align/determine beam pattern of Set A/Set B, including RS transmission to align with Rx beam assumption at UE side*  *• Indication of the preferred size of training dataset*  *• Indication of the minimum periodicity of the RS transmission.* |
| Fujitsu[16] | *Proposal 3: For DL beam (pair) prediction with a UE-side model, study the potential specification impacts of data collection on*  * The request to NW about the required RSs of Set A*  * The association of Set B and Set A*  * FFS on the method, e.g., the identifier of RS set or a bitmap to set the association of Set B and Set A.* |
| CMCC[17] | *Proposal 6: For data collection of DL Tx beam prediction at UE side, the Rx beam assumption should be aligned between the network and UE.*  *Proposal 7: For data collection of a UE-side AI/ML model, study the following options for association/mapping of beams within Set A and beams within Set B which is indicated from NW to UE:*  *· Opt1: QCL information between beams within Set A and beams within Set B*  *· Opt2: Identifier of info representing the association/mapping of Set A and Set B*  *e.g. start and ending indicator*  *· Opt3: The bitmap or pre-defined rule that beams of Set B is subset of beams of Set A* |
| Nokia[19] | *Proposal 17. For UE-sided BM-Case1 with DL Tx-Rx beam pair prediction, consider methods to reduce the necessary measurement space for DL TX-RX beam pair prediction at the UE side.*  *Proposal 18. For UE-sided BM-Case1 with DL Tx-Rx beam pair prediction, consider methods to indicate the necessary measurement space of preferred DL RS beam at the UE side.*  *Proposal 19. On data collection for model training for BM-case1 and BM-case2, enhancements for having dedicated RS measurements or reporting framework are not considered for model training.*  *Proposal 20. On data collection for model monitoring, Option 1 is supported.*  *• Option 1: data collection initiated/triggered by configuration from NW*  *Proposal 21. For UE-sided BM-case1 and UE-sided BM-case2, for functionalities supported towards the UE, RAN1 shall consider the required CSI-RS measurement enhancements for data collection at the UE side.*  *• Allowing the measurements of Full or partial Set A (associated with a functionality) beam measurements with a longer periodicity than the Set B measurements can be considered.* |
| Apple[20] | *Observation 2a: It is likely different AI/ML BM models need to be trained for different higher layer (dataset) IDs.*  *Observation 2b: If dissimilar configurations/scenarios can be assigned to the same higher ID, the UE side lacks a means to determine the composition of its data collection matches the overall deployment.*  *Observation 2c: more studies are needed to establish the feasibility for UE side training/inference.* |
| Xiaomi[21] | *Proposal 15: For data collection for AI/ML model training at UE side, support gNB indicating the relationship between set B and set A.* |
| OPPO[23] | *Observation 4: For the simplest case of BM-Case1 and BM-Case2 with UE-side AI/ML model training, data set including model inputs and labels can be collected by UE via legacy approach.*  *Proposal 2: For UE-side model training, it is up to UE to request the data collection followed by configuration of Set A and/or Set B from NW.* |
| Samsung[24] | *Proposal 12. For BM-Case1 with a UE-side AI/ML model, for data collection, support the configuration of spatial domain information of Set A and/or Set B, where identifiers can be used for representing Set A beams.*  * the spatial domain information of Set A and/or Set B should not disclose network implementation*  *Proposal 13. For BM-Case1 with a UE-side AI/ML model, for data collection, at least the following are identified with potential specification impact:*  * UE provides/requests for supported/preferred configurations of DL RS transmission (corresponding to Set A and/or Set B)*  * gNB provides the configurations associated with Set A and/or Set B*  * The initiation/triggering of data collection (e.g., by UE or gNB)*  *Proposal 27. For BM-Case2 with a UE-side AI/ML model, for data collection, at least the following information of data with potential specification impact are identified on the top of BM-Case1:*  * Request from UE for supported/preferred time domain configurations of DL RS transmission (corresponding to Set A and/or Set B).* |
| Lenovo[27] | *Proposal 7: Study data collection procedure to support both UE-side and NW-side AI/ML model training and model update (if supported).*  * For UE-side model training, study to support UE triggered or UE initiated data collection procedure,*  * For NW-side model training, support to report larger number of beams in one beam report by considering beam report via MAC CE or RRC.* |
| KT[29] | *Proposal 2. For model training at UE side, the initiation/trigger of the data collection by configuration from NW should be based on additional information or request from a UE.*  *Proposal 3. Consider how to stop/terminate the data collection for model training at UE side.*  *Proposal 4. For the model training/inference/monitoring of the UE-side AI/ML model:*  *• Study how to identify the beams within Set A which is not measured.*  *• Study how to indicate the association/mapping of beams within Set A and beams within Set B.* |
| QC[30] | *Proposal 2: For BM-Case1 and BM-Case2 with a UE-side AI/ML model, regarding aspects related to association/mapping of beams within Set A and beams within Set B, study mechanisms to ensure consistency of Set B beams and Set A beams across training and inference, including one or more of the following*  *• Set size consistency for Set B, Set A: consistency in number of beams and/or associated resources for Set B and Set A, across training and inference*  *• Order/indexing consistency: consistency in ordering of resources (e.g., resource index consistency) for Set B beams and Set A beams, across training and inference*  *• QCL consistency: consistency in QCL relationship of Set A beams with respect to Set B beams, across training and inference*  *• Beam shape consistency: relative pointing direction and beamwidth difference between physical beams with respect to Set A and Set B resources across training and inference should be under predefined tolerances* |
| MTK[31] | *Proposal 5: study the potential specification impact of UE reporting to network the supported/preferred configurations of DL RS transmission through reporting the following aspect:*  *• the supported Set A and Set B of beams*  *• the supported observation/prediction window sizes for BM-Case2*  *• the supported Rx beam assumptions*  *Proposal 6: Study the spec impact of configuring Option1 (data collection initiated/triggered by configuration from NW) by the CSI measurement and report configurations.*  *Proposal 7: Option2 (request from UE for data collection) can be studied together with UE-side performance monitoring procedures.* |
|  |  |

###### **Mod’s assessment**

* **Mod’s assessment:** The proposals in tdocs are quite diverging. Each detailed solution seems only proposed by very limited number companies (e.g., 1 or 2 companies). Moreover, many proposals seem not adding much no information compared to the existing agreements.
* **Mod’s suggestion:** No discussion until some proposal(s) is suggested by more companies

|  |  |
| --- | --- |
| Company | Comments |
| QC | We have extensively explained why we need to have some consistency between Set A and Set B beams across training and inference in our Tdoc. This is a critical issue that deserves discussion and debate. We suggest FL draft a proposal on this to trigger discussions. |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

###### Proposal 2.3.1 (FL4)

***~~Proposal 2.3.1: For BM-Case1 and BM-Case2 with a UE-side AI/ML model, regarding aspects related to association/mapping of beams within Set A and beams within Set B, study mechanisms to ensure consistency of Set B beams and Set A beams across training and inference, including one or more of the following~~***

* ***~~Order/indexing consistency: consistency in ordering of resources (e.g., resource index consistency) for Set B beams and Set A beams, across training and inference~~***

***Proposal 2.3.1:***

***Conclusion:***

***For BM-Case1 and BM-Case2 with a UE-side AI/ML model, regarding aspects related to association/mapping of beams within Set A and beams within Set B, conclude thatmechanisms to ensure consistency of Set B beams and Set A beams across training and inference (such as order/indexing consistently) should be considered.***

|  |  |
| --- | --- |
| Company | Comments |
| Mod | As requested by QC, LGE, ZTE, MTK, CATT, a new proposal is captured to collect companies’ view |
| ZTE | Fine to discuss it or in the WI phase. As evaluated in agenda 9.2.3.1, Set B of beams can be fixed, changed among multiple pre-configured beam patterns, or randomly changed among Set A of beams (if supported). If Set B is fixed, the legacy mechanism can be reused to guarantee the indexing consistency. However, if Set B can be changed among multiple pre-configured beam patterns or randomly changed among Set A of beams, the legacy mechanism may not be sufficient depending on the generalization ability of the trained UE-side model. |
| HW/HiSi | Not needed at this stage.  We already have agreements to ensure that these aspects are taken into account (please see below the mapping between Set A and Set B). In our view, a potential issue or problem will not occur. And if any issue would be identified during a potential WI, the detail to overcome it would be dependent on the outcome of other related still ongoing discussions and if needed should be solved then.    Agreement  For BM-Case1 and BM-Case2 with a UE-side AI/ML model, study potential specification impact of AI model inference from the following additional aspects on top of previous agreements:   * Indication of the associated Set A from network to UE, e.g., association/mapping of beams within Set A and beams within Set B if applicable * Beam indication from network for UE reception * Note: The second bullet may or may not have additional specification impact (e.g., legacy mechanism may be reused).   Agreement  Regarding the data collection for AI/ML model training at UE side, study the potential specification impact considering the following additional aspects.   * Whether and how to initiate data collection * Configurations, e.g., configuration related to set A and/or Set B, information on association/mapping of Set A and Set B * Assistance information from Network to UE (If supported) * Other aspect(s) is not precluded |
| CMCC | Consistency in ordering of resources for Set A beams across training and inference can be guaranteed by configuring RS ID of set A beams in pre-defined order. Consistency in ordering of resources for Set B beams across training and inference can be guaranteed by configuring association/mapping of Set A and Set B, which is included in previous agreement. |
| Fujitsu | We are fine for the main bullet. In legacy method which beam carries the configured RS depends on the gNB implementation. But UE-side model may have requirement on gNB behavior about the consistency on the mapping Tx beams and RS during training and inference. The Order/indexing consistency may just be one of aspects of consistency. |
| CATT | We are fine with the intention of this proposal. But we ‘d like to remove the e.g. in the bracket. Since this is still an open issue and need more time to consider and discussion. At this stage, the ordering of resource is sufficient for order/indexing consistency of set a and set B.  ***Proposal 2.3.1: For BM-Case1 and BM-Case2 with a UE-side AI/ML model, regarding aspects related to association/mapping of beams within Set A and beams within Set B, study mechanisms to ensure consistency of Set B beams and Set A beams across training and inference, including ~~one or more of~~ the following***   * ***Order/indexing consistency: consistency in ordering of resources for Set B beams and Set A beams, across training and inference*** |
| Ericsson | We would like to avoid proposals on “what to study” at this stage. In general for ML, there would be a performance degradation if you perform inference on other data than what you trained on. Hence, we can agree to the intention of the proposal, we would however avoid examples, since there would be multiple alternatives that can enable such consistency.  ***Proposal 2.3.1 for Conclusion: For BM-Case1 and BM-Case2 with a UE-side AI/ML model, regarding aspects related to association/mapping of beams within Set A and beams within Set B, ~~study~~ conclude thatmechanisms to ensure consistency of Set B beams and Set A beams across training and inference can be beneficial to avoid performance degradation~~, including one or more of the following~~***   * ***~~Order/indexing consistency: consistency in ordering of resources (e.g., resource index consistency) for Set B beams and Set A beams, across training and inference~~*** |
| Mod | The proposal is updated. Please continue to input comments (if any) |
|  |  |
|  |  |

# Spec impact of AI/ML inference for BM-Case1 & BM-Case2

## General/common aspects

In previous RAN1 meeting(s), the agreement(s)/conclusion(s) were made as below:

|  |
| --- |
| **RAN1#110**  Agreement  In order to facilitate the AI/ML model inference, study the following aspects as a starting point:   * Enhanced or new configurations/UE reporting/UE measurement, e.g., Enhanced or new beam measurement and/or beam reporting * Enhanced or new signaling for measurement configuration/triggering * Signaling of assistance information (if applicable) * Other aspect(s) is not precluded   Agreement  For the sub use case BM-Case1 and BM-Case2, further study the following alternatives for the predicted beams:   * Alt.1: DL Tx beam prediction * Alt.2: DL Rx beam prediction * Alt.3: Beam pair prediction (a beam pair consists of a DL Tx beam and a corresponding DL Rx beam) * Note1: DL Rx beam prediction may or may not have spec impact   **RAN1#112**  Conclusion  For the sub use case BM-Case1 and BM-Case2, “Alt.2: DL Rx beam prediction” is deprioritized.  **RAN1#113**  Agreement  For BM-Case1 and BM-Case2, study necessity, benefit(s) and potential specification impact from the following additional aspects for AI model inference:   * How to perform beam indication of beams in Set A not in Set B.   + Note: the legacy mechanism may be sufficient. |

The related proposals in tdocs are copied as below:

|  |  |
| --- | --- |
| NVIDIA[3] | *Proposal 8: For AI/ML based beam prediction in spatial/time domain, study potential specification impact related to report/feedback of model input for inference, type of model input, and model input acquisition and pre-processing.*  *Proposal 9: For AI/ML based beam prediction in spatial/time domain, study potential specification impact related to report/feedback of model inference output and post-processing.* |
| Huawei[4] | *Proposal 13: For BM-Case1 and BM-Case2 the legacy mechanism is sufficient for beam indication of beams in Set A not in Set B.* |
| IDC[6] | *Observation 20: For Rel-15 beam management, actual mapping between DL Tx beam and UE Rx beam is totally based on UE implementation.*  *Observation 21: The implementation-based UE Rx beam selection works for Rel-15, however, UE Rx beam information is crucial to accurately predict beam qualities for AI/ML based beam prediction.*  *Proposal 33: Study benefits of specification enhancements on acquiring UE Rx beam information for DL Tx beam prediction (Alt. 1) and beam pair prediction (Alt. 3).* |
| ZTE[5] | *Observation 7: The legacy mechanism is sufficient to perform* *beam indication of beams in Set A but not in Set B.*  *Observation 8: An additional aperiodic RS resource set can be triggered for beam measurement over the predicted top-1/K beams for the purpose of beam indication of beams in Set A but not in Set B, which is up to NW implementation.*  *Proposal 20: In order to facilitate AI/ML operations for BM-Case1 and BM-Case2, study the following additional aspects:*  *• Beam indication of the predicted DL beam pair from network to UE*  *• Beam indication of multiple future time instances in one indication for BM-Case2* |
| Vivo[7] | *Proposal 33: For BM-Case1 and BM-Case2 with a network-side AI/ML model, study potential specification impact on TCI indication for AI/ML model inference:*  *• Enhanced TCI indication based on both Rel-15/16 and Rel-17 unified TCI frameworks*  *Proposal 38: For BM-Case1 and BM-Case2 with a UE-side AI/ML model, study potential specification impact on TCI indication for AI/ML model inference:*  *• Enhanced TCI indication based on both Rel-15/16 and Rel-17 unified TCI frameworks* |
| Ericsson[11] | *Observation 5 In most scenarios, there are no issues of beam indication for PDSCH transmission since UE needs to:*  *a. perform measurements on the Top-K beams to find the strongest beam,*  *b. perform measurements to estimate the signal strength for link adaptation purpose also for K=1*  *c. perform P3 sweep also for K=1*  *Proposal 8 Conclude that there is no issue of beam indication for scenarios where the UE measures Top-K beams prior to PDSCH transmission when K>1* |
| Google[12] | *Proposal 15: For AI/ML based BM, the study should be based on both Rel-17 unified TCI framework and Rel-15/Rel-16 BM framework.* |
| LG[13] | *Proposal #3: It should be assumed that measurements of Set A beams are not available or is available but can be outdated at UE side, and UE is not expected to manage a good Rx beam for each of the Set A beams.*  *Proposal #9: For beam indication, TCI/QCL RS should be based on Set B beams of which UE can measure and maintain its Rx beam. To indicate one beam in Set A not in Set B, consider indicating multiple neighboring beams from Set B for helping UE to find its Rx beam for the Set A beam.* |
| CATT[14] | *Proposal 10: For BM-Case1 and BM-Case2, for beam indication of the predicted Top-K beams, the following options can be considered:*  *• Opt1: Reusing legacy TCI indication mechanism after sweeping over the Top-K predicted beams*  *• Opt2: Indicating a measured beam close to the predicted best beam*  *Proposal 11: For BM-Case2, indicating beams for multiple future time instances will limit the beam indication flexibility of the network.* |
| Nokia[19] | *Observation 9. If DL Tx beam prediction is applied for BM-Case2, specification impact on the current TCI beam indication framework while BM-Case2 with the possibility of multiple beams indication is applied, should be discussed.*  *Observation 10. If DL Tx beam prediction is applied for BM-Case2, changing indication from one TCI state to another one imposes a considerable overhead even though the entire predicted beams within the indication process come from the same prediction procedure specifically if a shorter prediction window size (e.g., 40ms) is applied,*  *Proposal 26. RAN1 to support enhancements to the Rel-17 unified TCI state beam indication framework aiming at beam indication overhead reduction.*  *• Support multi beams (sequence of beams)/ multi TCI states via single beam indication to minimize the ignaling overhead.* |
| DCM[22] | *Proposal 3: Study how to indicate the beam of Set A that is not included in Set B/C. At least the following approaches can be considered.*  *・Beam indication of Set A via QCL source of Set B/C*  *・Beam indication of Set A via QCL source of Set A with additional measurement of Set A* |
| MTK[31] | *Proposal 20: Deprioritize the study of performing beam indication of beams in Set A not in Set B*  *Proposal 21: Considering performing beam indication of beams in Set A not in Set B by legacy mechanism where NW configures additional DL RS for UE measurement.* |
|  |  |

###### Mod’s assessment

**Beam indication of beams in Set A not in Set B**

At least 4 companies (e.g., ZTE, Ericsson, CATT, MTK) declare that legacy mechanism is sufficient to perform beam indication of beams in Set A not in Set B, whereas there is one company (LG) proposes some new mechanism.

* **Mod’s assessment:** In that last meeting, the group agreed to study “How to perform beam indication of beams in Set A not in Set B.” and a note “the legacy mechanism may be sufficient” was included to reflect the potential study output of no spec impact. Based on previous discussions, many other companies thought there is no issue. Thus, it is better to make some conclusion on this issue
* **Mod’s suggestion:** A proposal is suggested for further discussion.
* **Related proposals in tdocs**
  + ZTE: Observation 7
  + Ericsson: Proposal 8
  + CATT: Proposal 10
  + MTK: Proposal 20, 21
  + Huawei: Proposal 13
  + LG: Proposal 9

**Use R15/R16/R17 signalling framework**

At least 2 companies (e.g., vivo, Google) are discussing whether the signaling framework of AI-based BM should be R15/16 signaling framework and/or R17/18 unified TCI framework.

* **Mod’s assessment:** This topic seems more suitable for the discussion in work item.
* **Mod’s suggestion:** No discussion in this meeting.
* **Related proposals in tdocs**
  + Vivo: Proposal 33, 38
  + Google: Proposal 15
  + DCM: Proposal

###### Proposal 3.2.1 Closed

* Legacy mechanism
  + ZTE: Observation 7
  + Ericsson: Proposal 8
  + CATT: Proposal 10
  + MTK: Proposal 20, 21
* New mechanism
  + LG: Proposal 3,9
  + DCM: Proposal

***Proposal 3.2.1:***

***Conclusion***

***For BM-Case1 and BM-Case2 with a UE-side AI/ML model, for AI model inference, the legacy mechanism can be used to perform beam indication of beams in Set A not in Set B***

* ***No consensus on the need of any specification enhancement from RAN1 perspective***

|  |  |
| --- | --- |
| Company | Comments |
| LG | Not support. From technical perspective, we have concern on the argument such that legacy TCI indication works well since it works with ‘unknown’ TCI state as well as ‘known’ TCI. It is true but if Set B beams are ‘known’ and those Set A beams not in Set B are ‘unknown’, why don’t we try to use ‘known’ TCI for beam indication to save measurement/RS overhead and latency to avoid RX beam refinement procedure?  In general, since this is study item and we agreed this study point at the last meeting, we suggest to leave the door open for any enhancement techniques for Rel-19 WI (if supported). This does not mean that we shall support any enhancement which will be up to WI discussion/decision. |
| Xiaomi | For UE-side model, from our pint of view, if UE can report this beam ID, legacy beam indication mechanism can be reused. |
| NTT DOCOMO | Does the legacy mechanism include the Set A beam indication based on Set B/C as QCL source, even when Set A is not in Set B? Anyway, this point should be clarified in the proposal. |
| CATT | We are fine with the proposal since the late stage for the SI. At least the second-stage Top-K beam sweeping with legacy beam indication can work. |
| ZTE | Support. According to clause 8.10 of TS 38.133, the configured TCI state can be known or unknown as defined below and different TCI switching delay will be applied accordingly. Therefore, the legacy mechanism is sufficient to perform beam indication of unknown beams in Set A but not in Set B. Additionally, the beam in the measured beam set can be indicated or an additional aperiodic RS resource set can be triggered for beam measurement over the predicted top-1/K beams to obtain the corresponding QCL relationship, which is up to NW implementation.   |  | | --- | | The TCI state is known if the following conditions are met:  - During the period from the last transmission of the RS resource used for the L1-RSRP measurement reporting for the target TCI state 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  - The UE has sent at least 1 L1-RSRP 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 ≥ -3dB  Otherwise, the TCI state is unknown. | |
| Samsung | Fine with the intention. This can be a conclusion rather than proposal. |
| HW/HiSi | Support |
| QC | We had agreed to study potential spec impact for beam indication, but it is not clear to us why we should close the door for potential related enhancements at this point, following similar logic as LG. |
| Lenovo | If the beam indication in the main bullet refers to the TCI state or QCL indication for DL reception or UL transmission, we agree that legacy mechanism may be enough. It seems it should be a conclusion 3.1.1 other than a proposal. |
| Futurewei | Support. We think it is OK to use legacy mechanism for model inference. |
| Ericsson | Support. Don’t understand the concern by LG. We agreed to study it in last meeting, and based on our study, we don’t see a need for spec. enhancements. |
| Nokia/NSB | OK |
| Spreadtrum | Support |
| Sony | Share with QC’s view, we think it can be left for further studying the enhancements and potential spec impact at this stage |
| Mod | * 10 companies support * 3 companies not support   @LG, QC, Sony: Maybe the proponent can share more details of the solution in your mind so that other companies can be convinced |
| MediaTek | We are OK with this proposal. |
| CMCC | If beam sweeping of beams in Set A not in Set B is needed, legacy TCI mechanism works well. |
| Fujitsu | If the beam indication in main bullet means the TCI state or QCL indication, we are fine to this proposal. |
| Google | We do not quite understand the intention for the proposal. Regarding beam indication, we think one important aspect is to decide whether further study is based on Rel-16 framework or Rel-17 framework or both. |
| Xiaomi | Fine with the conclusion |
| NEC | Support |
| HW/HiSi | Support |
| Mod | Summary of companies’ view   * 14 companies support * 4 companies not support (LGE, Sony, Apple, QC ) |
| Mod | Closed. Please see the corresponding agreement in Chair’s note |

## AL/ML inference at gNB side

In previous RAN1 meeting(s), the agreement(s)/conclusion(s) were made as below:

|  |
| --- |
| **RAN1#110bis-e**  Working Assumption  For BM-Case1 and BM-Case2 with a network-side AI/ML model, study the following L1 beam reporting enhancement for AI/ML model inference   * UE to report the measurement results of more than 4 beams in one reporting instance * Other L1 reporting enhancements can be considered   **RAN1#111**  Agreement  For BM-Case1 and BM-Case2 with a network-side AI/ML model, study potential specification impact on the following L1 reporting enhancement for AI/ML model inference   * UE to report the measurement results of more than 4 beams in one reporting instance * Other L1 reporting enhancements can be considered   **RAN1#113**  Agreement  For BM-Case2, study necessity, benefit(s) and potential specification impact from the following additional aspects for AI model inference:   * Reporting information about measurements of multiple past time instances in one reporting instance for BM-Case2   + Note: only applicable to network-side AI/ML model * Note: The potential performance gains of measurement reporting should be justified by considering UCI payload overhead |

The related proposals in tdocs are copied as below:

|  |  |
| --- | --- |
| H3C[1] | *Proposal 1: For BM-Case1 and BM-Case2 with NW-side model, further study Alt.1 DL Tx beam prediction with high priority.* |
| Huawei[4] | *Observation 2: For the beam prediction mechanisms for BM-Case1 and BM-Case2, Alt.1 (DL Tx beam prediction) is a natural replacement of the legacy P1/P2 procedure for Tx beam sweeping, and is compatible with any pattern of the Rx beams.*  *Observation 3: For DL Tx beam prediction when the AI/ML model is at the network-side, there is no need to introduce additional types of information other than the report of CRI/RSRP, etc., which is already supported by legacy releases.*  *Observation 4: For the AI/ML-based DL Tx beam prediction, non-AI/ML options can be implemented to optimize the Rx beam selection*  *• Opt.1: Fixed Rx beams is used for inference during P1/P2 and the Rx beam sweeping is performed to determine the Rx beam in P3*  *• Opt.2: A quasi-optimal DL Rx beam can be identified by sweeping the always-on SSB beams at P1 and used for Tx beam prediction at P2*  *• Opt.3: Exhaustive Rx beam sweeping is swept over multiple P1/P2 periods each of which predicts the best Tx beam for a specific Rx beam*  *Observation 5: DL Tx-Rx beam pair prediction is much more complicated to evaluate and a thorough assessment has not yet been performed. At least following issues have to be be taken into account to assess its performance:*  *• UE rotations and Rx beam blocking (when applicable)*  *• RSRP measurement errors*  *• Performance/overhead/latency*  *• Complexity*  *Proposal 7: In Rel-18 SI, not to further discuss the spec impact for DL Tx-Rx beam pair prediction on top of DL Tx beam prediction.*  *Observation 10: For the inference of BM-Case2, the motivation for indicating Top-1/K Beam of multiple future time instances in one-shot manner is not clear.*  *Proposal 16: To save UE reporting overhead for the inference of the AI/ML model at the network side, Opt.1 (M1 L1-RSRPs and indication) for data collection can be reused so that UE can feedback the RSRP values for a subset of all measured beams in Set B.*  *Proposal 17: For AI/ML model at the network side, no strong motivation to introduce finer resolution for UE reported measurement results at least for model inference.* |
| Spreadtrum[5] | *Proposal 2: For sub use cases BM-Case1 and BM-Case2, support both Alt 1 and Alt 3.*  *Observation 1: For beam pair prediction in BM-Case1 and BM-Case2*  *- If AI/ML inference is at NW side, the indication of Rx beam for unmeasured beam should be enhanced.*  *Proposal 7: For beam pair prediction with a network-side AI/ML model, study the enhancement for beam reporting to report one DL Tx beam received by multiple Rx beams.*  *Proposal 8: For BM-Case1 and BM-Case2 with a network-side AI/ML model, existing quantitative criteria should be reused.*  *Proposal 9:* *Reporting multiple past time instances in one reporting instance for BM-Case2 is not needed.*  *Proposal 10: For sub use cases BM-Case2, implicit indication or report of time information should be considered.* |
| IDC[6] | *Proposal 2: For BM-Case 2 with a gNB-side AIML model, do not support reporting measurements of multiple past time instances in one report.* |
| Vivo[7] | *Proposal 13: Support expected Rx beam information as the AI input as one of the solutions on NW-side beam prediction for generalization to different number of Rx beams.*  *Proposal 14: Support to further study specification impact on NW-side beam pair prediction. Consider to train sufficient number of UE locations and orientations to address the coordination system mismatch issue.*  *Proposal 28: For BM-Case1 and BM-Case2 with a network-side AI/ML model, study potential specification impact on resource configuration for AI/ML model inference:*  *• Specific beam pair resource configuration for Set B*  *• Enhanced P3+P2 resource configuration that Rx beam assumption of P2 resource measurement is the best Rx beam searched from P3 procedure for Set B*  *• Resource configuration for Set C*  *Proposal 30: For BM-Case1 and BM-Case2 with a network-side AI/ML model, study potential specification impact on measurement report for AI/ML model inference:*  *• UE measures the beams of Set B/Set C and reports M1 L1-RSRPs optionally with M2 RS indicators*  *- If M1 is equal to the number of beams or beam pairs in Set B (noted as X), corresponding RS indicators may be not needed.*  *- If M1 is smaller than X/2, corresponding M2 RS indicators are needed*  *- If M1 is smaller than X, but larger than X/2, RS indicators are needed for indicating M2 beams or beam pairs in Set B not included in the measurement report.*  *Proposal 31: For BM-Case1 and BM-Case2 with a network-side AI/ML model, study potential specification impact on report overhead reduction for AI/ML model inference:*  *• Reducing unnecessary L1-RSRP report where the omitted L1-RSRPs may be lower than a pre-defined threshold*  *• Pattern-based beam report if beam resource configuration with multiple pre-configured patterns is supported*  *• Study how to further reduce report overhead of time domain beam prediction for measurement results of multiple occasions.*  *Proposal 32: For BM-Case1 and BM-Case2 with a network-side AI/ML model, study potential specification impact on quantization enhancement for RSRP quality improvement for AI/ML model inference:*  *• High-precision L1-RSRP quantization*  *• Multi-resolution L1-RSRP quantization, e.g. high-resolution quantization for a group of best RSRPs and low-resolution quantization for others.* |
| ZTE[8] | *Proposal 10: For BM-Case1 and BM-Case2 with a network-side AI/ML model, study the following L1 beam reporting enhancements for AI/ML model inference:*  *• Reporting resolution enhancement*  *• Reporting overhead reduction*  *• Reporting assistance information*  *• Reporting of multiple past time instances in one reporting instance* |
| Intel[9] | *Proposal 1: Beam Pair prediction (Alt-3) should be supported, at least for BM-Case 1 since it can provide large latency and measurement gains for joint P2/P3 procedure*  *Observation 2: For beam pair prediction at network side using DL measurements, to align reports from multiple UEs, the network may configure an abstract framework related to spherical coverage through explicit or implicit indication and the UEs can measure DL signals on beams corresponding to configured indexes and report the measurements to gNB which can then construct set B.* |
| Ericsson[11] | *Proposal 10 For BM-Case1 and BM-Case2 with a network-side AI/ML model, study the following additional aspects (including the necessity) to facilitate AI model inference:*  *• For BM-case 2. Report of compressed value(s) based on temporal sequence of L1-RSRP (e.g. temporal variance or polynomial approximation of L1-RSRP/L1-SINR measurements for beams)* |
| Google[12] | *Proposal 9: Support the NW configures a measurement window for the UE to identify one beam quality from multiple SSB/CSI-RS instances and configures the UE to report the beam quality for an SSB corresponding to multiple measurement windows.* |
| LG[13] | *Proposal #10: For NW-sided AI/ML in BM-Case2, consider the following UE reporting enhancements*  *- past/present best N beam(s) per time stamp*  *- tendency/variance of best N beam(s)*  *Proposal #13: For NW-sided model, Tx beam prediction should only be considered.* |
| NEC[15] | *Proposal 10: Regarding model inference for BM-Case2 with NW-side AI/ML model, study discontinuous P/SP beam report.* |
| Fujitsu[16] | *Proposal 4: Regarding the inference of NW-side AI/ML model, study the potential specification impacts on the UE behavior of beam reporting.*  *Proposal 5: For the DL beam pair prediction with a NW-side AI/ML model, study the potential specification impacts on the Rx beam information included in report instance.*  *Proposal 6: Regarding the Rx beam information included in report instance for the DL beam pair prediction with a NW-side AI/ML model, it is suggested to study.*  * Physical beam information (e.g., beam angle)*  * Logical beam information (e.g., beam ID)*  * FFS: How to map the logical beam with the physical beam* |
| CMCC[17] | *Proposal 1: Regarding beam prediction type, Tx beam prediction at NW side is prioritized, Tx beam/beam pair prediction at UE side can be considered, beam pair prediction at NW side is not considered.*  *Proposal 8: For BM-Case1 with a network-side AI/ML model, study the following L1 beam reporting enhancement for AI/ML model inference*  *• whether Rx beam related information corresponding to a measured Tx beam reported from UE to NW*  *• How to indicate association/mapping of beams within Set A and beams within Set B from NW to UE*  *· Opt1: QCL information between beams within Set A and beams within Set B*  *· Opt2: Identifier of info representing the association/mapping of Set A and Set B*  *e.g. start and ending indicator*  *· Opt3: The bitmap or pre-defined rule that beams of Set B is subset of beams of Set A*  *Proposal 9: For BM-Case1 with a network-side AI/ML model, study the acknowledgement mechanism of available Rx beams for AI/ML model inference.*  *Proposal 16: For BM-Case2 with a network-side AI/ML model, study feasibility, necessity, benefit(s) and potential specification impact from the following additional aspects for AI model inference:*  *• Beam indication of multiple future time instances for BM-Case2*  *- FFS: applicable for Top-1 and/or Top-K predicted beams*  *• Overhead reduction for the reporting of L1-RSRP measurement results*  *- FFS: e.g. reporting a partial Set B, L1-RSRP quantization, compressed temporal information for BM-Case2, statistics of past measurements for BM-Case2, etc.* |
| Nokia[19] | *Proposal 9. For NW-sided BM-Case1, the following potential specification impact can be considered,*  *• For model inference at the NW, enhancements to the CSI reporting such that the UE can be configured to report the measurement results of more than 4 beams and corresponding L1-RSRP in one beam reporting instance.*  *• For beam indication/activation towards the UE, enhancements to the CSI reporting to enable beam measurement and reporting of beams corresponding to the Top-K predicted beams.*  *• For performance monitoring at the NW, enhancements to the CSI reporting may not be needed to enable full/partial Set A beam measurements.*  *Proposal 10. For NW-sided BM-Case2, the following potential specification impact can be considered,*  *• For model inference at the NW, enhancements to the CSI reporting such that the UE can be configured to report the measurement results of more than 4 beams and corresponding L1-RSRP in one beam reporting instance.*  *• For model inference at the NW, enhancements to the CSI measurement and reporting such that the UE can be configured to measure DL RS and report the measurement results for a T1 duration of time and deactivate the measurements/reporting for a T2 duration of time.*  *• For beam indication/activation towards the UE, during T2 duration of time, enhancements to the CSI reporting to enable beam measurement and reporting of beams corresponding to the Top-K predicted beams.*  *• For performance monitoring at the NW, enhancements to the CSI reporting may not be needed to enable full/partial Set A beam measurements.*  *Observation 4: For NW-sided DL beam pair prediction of BM-Case1, information about the Rx beam codebook needs to be coordinated with NW, either by indicating an ID corresponding to a shared dataset or information about the Rx beam ID at coarse quantized angles.*  *Proposal 11. For BM-Case1, consider the following,*  *• For the construction of Set A/B, prioritize Alt.2: Set B is a subset of Set A.*  *• For beam types of Set A/B, prioritize Alt.1: DL Tx beam prediction.*  *Proposal 12. For BM-Case2, consider the following,*  *• For the construction of Set A/B, prioritized “Set B and Set A are the same”.*  *• For the beam types of Set A/B, prioritize Alt.1: DL Tx beam prediction.* |
| Apple[20] | *Proposal 1: Deprioritize beam pair prediction in the study.*  *Proposal 5: capture 4 beam reporting methods for BM Case-1 and BM Case-2:*   |  |  |  |  |  | | --- | --- | --- | --- | --- | | BM Case-1 | | | | | | S = # of set A beams, M = # of set B beams, N = # of selected B beams for beam reporting | | | | Example (S=32, M=16, N=10) | |  | Beam selection | RSRP | Total signaling bits | | Reporting all set B beams’ RSRPs | for the strongest beam |  |  | 71 | | Reporting selected set B beams’ RSRPs with CRI signaling |  |  |  | 83 | | Reporting selected set B beams’ RSRPs with bitmap |  |  |  | 63 | | Reporting selected set B beams’ RSRPs with combinatorial index |  |  |  | 60 |  |  |  |  |  |  | | --- | --- | --- | --- | --- | | BM Case-2 | | | |  | | S = # of set A beams, M = # of set B beams, N = # of selected B beams for beam reporting, W = # of observation intervals | | | | Example (S=32, M=16, N=10, W=4) | |  | Beam selection | RSRP | Total signaling bits | | Reporting all set B beams’ RSRPs | for the strongest beam over W intervals |  |  | 265 | | Reporting selected set B beams’ RSRPs | (strongest beam indication per interval) | per interval |  | 332 | | Reporting selected set B beams’ RSRPs with bitmaps | per interval | per interval |  | 252 | | Reporting selected set B beams’ RSRPs with combinatorial index | per interval | per interval |  | 240 | |
| Xiaomi[21] | *Proposal 1: Support L1-RSRP and beam (pair) ID as AI/ML model input with high priority for variable set B.*  *Proposal 2: For spatial domain beam pair prediction, consider to report Rx beam information, including Rx beam ID of UE to gNB for gNB side inference.*  *Proposal 3: For beam indication of Tx beam being not measured by UE, gNB can indicate the Rx beam ID instead of Tx beam ID to UE in the case of Tx/ Rx beam pair prediction at gNB side.*  *Proposal 4: For the case of Tx beam or TxRx beam pair inference with specific Rx, support to indicate Rx beam information to UE for obtaining L1-RSRP input to AI/ML model.*  *Proposal 7: Support to report predicted L1-RSRP in the L1-beam report with an indication to let gNB know which L1-RSRP is a predicted L1-RSRP.* |
| DCM[22] | *Observation 4: In DL Tx-Rx beam pair prediction with NW side model, some mechanisms to report Rx beam ID used for beam measurements are necessary.*  *Proposal 6: If RAN1 can make the consensus that the DL Rx beam information cannot be reported to NW, DL Tx-Rx beam pair prediction with NW side model should be deprioritized due to the feasibility.*  *Proposal 7: Study the potential specification impacts of Rx beam determination policy for measurement reporting for NW side beam prediction. E.g., the measurements results from the same Rx beam are reported for Set B.*  *Proposal 9: Study at least the following UCI payload overhead reduction via L1 beam reporting of multiple measurement time instances in one reporting instance for temporal beam prediction.*  *・Common CRI/SSBRI at multiple time instances to track the beam quality*  *・Differential L1-RSRP value representation from ones at different time instances.* |
| OPPO[23] | *Proposal 12: For BM-Case1 with NW-side model, study whether/how to reduce the reporting overhead of both fixed or variable Set B, e.g. by dropping the part of SSBRIs/CRIs.*  *Proposal 13: For BM-Case2 with NW-side model, further study how UE reports multiple instances of Set B measurements into one beam reporting instance.*  *Observation 11: For Tx beam prediction with NW-side model, it seems not necessary to enhance the signaling aspect, e.g. combining or associating the Tx beam indication and the DL Rx beam sweeping.*  *Proposal 14: For BM-Case1 and BM-Case2 with NW-side model, study the feasibility of Tx beam indication and/or beam pair indication.*  *Proposal 15: For BM-Case2 with NW-side model, study the feasibility of beam (pair) indication for multiple future time instance(s) in a single beam indication.*  *Proposal 22: For BM-Case1 and BM-Case2, support Tx beam prediction (Alt.1) and beam pair prediction (Alt.3).* |
| Samsung[24] | *Proposal 6. For BM-Case1 with a network-side AI/ML model, for the L1 signaling to facilitate model inference, the following at least the following with potential specification impact is identified:*  * Additional support of L1 beam report with the measurement results of more than 4 beams in one reporting instance.*  *Proposal 7: For BM-Case1 with a network-side AI/ML model, for model inference, the following aspects should be considered to support a single beam report with more than 4 beams in one reporting instance:*  * CSI report configuration*  * Content of CSI report*  * Payload size reduction*  *Proposal 8: For BM-Case1 with a network-side AI/ML model, study feasibility, necessity, benefit(s) and potential specification impact from the following additional aspects for AI model inference:*  * Spatial domain predictive beam indication for BM-Case1*  *Proposal 24. For BM-Case2 with a network-side AI/ML model, for the L1 signaling to facilitate model inference, at least the following with potential specification impact is identified on the top of BM-Case1:*  * Reporting information about measurements of multiple past time instances in one reporting instance for BM-Case2.*  *Proposal 25: For BM-Case2 with a network-side AI/ML model, at least the following with potential specification impact is identified:*  * A single beam indication for multiple future time instances* |
| ETRI[26] | *Proposal 2: For BM-Case2 with a NW-side AI/ML model, it can be considered to utilize the following compressed information to report measurements of multiple past time instances within a single reporting instance:*  *- Reporting the differential values from the measurement at a specific time instance.*  *- Reporting a statistical representative value, e.g. mean, median, maximum, etc.* |
| Lenovo[27] | *Proposal 1: Study the potential specification impact for the UE to report or obtain the RSRPs for a same Tx beam with different Rx beams for the AI/ML Model input for NW-side or UE-side AI/ML inference.*  *Proposal 2: Study schemes on differential RSRP report for UCI overhead reduction for larger number of beam report in a beam report.*  *Proposal 3: Study mechanism to adapt/change the Set B configuration or the beam selection for AI/ML model input for RS and UCI report overhead reduction.* |
| QC[30] | *Proposal 3: For BM-Case1 and BM-Case2 prioritize the study of DL Tx beam prediction.* |
| MTK[31] | *Proposal 1: For the sub use case BM-Case1 and BM-Case2, focus on Alt.1 (i.e., DL Tx beam prediction) and deprioritize Alt.3 for the predicted beams for further study.*  *Proposal 17: For the data collection for NW-side model for BM-Case2, it is not necessarily for UE to report information about measurements of multiple past time instances in one reporting instance for BM-Case2*  *Proposal 23: To facilitate AI model inference for a NW-side AI/ML model, study spec impact of reporting overhead reduction with quantizing L1-RSRP and/or normalized L1-RSRP measurement with lower number of bits than the current spec.* |
|  |  |
|  |  |

###### Mod’s assessment

**DL Tx beam prediction vs DL beam pair prediction**

Based on the tdocs and previous discussion, the preference of each companies seems not changed. Companies’ views in the tdocs are summarized as below

* Alt.1 (DL Tx beam prediction) is prioritized (at least for NW-side model)
  + Huawei, Ericsson, LGE, Nokia, Samsung, MTK, Apple, DCM, CMCC, QC, H3C
* Alt.3 (DL beam pair prediction) is preferred or support both Alt.1 and Alt.3
  + Futurewei, ZTE, Spreadtrum, OPPO, vivo, IDC, CATT, Fujitsu, CAICT, Intel, China Telecomm, Xiaomi, Huawei(support DL beam pair prediction for both NW-side and UE-side AI/ML model), Nokia (for UE-sided model) , H3C(for UE-side model)
* **Mod’s assessment:** This issue was raised from the 2nd RAN1 meeting and discussed at each previous meeting. However, by reading the tdocs, the situation seems not changed. Since there are many dedicated spec impacts needed for beam pair prediction and this meeting is the last one of R18 study item, the group have to make final decision/conclusion now. Based on tdocs and previous discussion, it seems unlikely to support DL beam pair prediction with the same priority compared to DL Tx beam prediction at least for NW-side AI/ML model. Meanwhile, some proponents of Alt.3 cannot accept any “negative statement” (e.g., deprioritized) on DL beam pair prediction. In order to move forward, some compromise is needed and all companies are encouraged to show constructiveness and flexibility.
* **Mod’s suggestion:** A compromised proposal including both UE-side and NW-side AI model is suggested for further discussion.
* **Related proposals in tdocs**
  + H3C: Proposal 1, 2
  + Huawei: Proposal 7
  + LG: Proposal 13
  + CMCC: Proposal 1
  + Nokia: Proposal 11, 12
  + Apple: Proposal 1
  + DCM: Proposal 7
  + QC: Proposal 3
  + MTK: Proposal 1
  + Spreadtrum: Proposal 2
  + Intel: Proposal 1
  + Fujitsu: Proposal 5
  + OPPO: Proposal 22

**Detailed design for DL beam pair prediction**

At least 4 companies (e.g., Spreadtrum, vivo, Fujitsu, Xiaomi) are discussing the detailed enhancement for DL beam pair prediction.

* **Mod’s assessment:** Due to the afore-mentioned background, it seems premature to discuss this issue at current stage
* **Mod’s suggestion:** Postpone the discussion until the group achieve consensus on the support of DL beam pair prediction.
* **Related proposals in tdocs**
  + Spreadtrum: Proposal 7
  + Vivo: Proposal 14, 28
  + Fujitsu: Proposal 5, 6
  + Xiaomi: Proposal 2, 3, 4

**Overhead reduction of L1 reporting**

At least 10 companies (e.g., vivo, ZTE, Ericsson, CMCC, DCM, OPPO, Samsung, ETRI, Lenovo, MTK) suggest some proposals for the overhead reduction of L1 reporting.

* **Mod’s assessment:** A proposal was suggested and discussed for overhead reduction of data collection. Although most companies support the proposal, no consensus was made since two companies commented that it was not needed to discuss this issue. As 10 companies suggest detailed solution for overhead reduction, it is an important issue from the views of many companies
* **Mod’s suggestion:** Discuss this issue in the data collection. No separate discussion in this session.
* **Related proposals in tdocs**
  + Vivo: Proposal 31
  + ZTE: Proposal 10
  + Ericsson: Proposal 10
  + CMCC: Proposal 16
  + DCM: Proposal 9
  + OPPO: Proposal 12
  + Samsung: Proposal 7
  + ETRI: Proposal 26
  + Lenovo: Proposal 2
  + MTK: Proposal 23

###### Proposal 3.2.1 Closed

***~~Proposal 3.2.1: For BM-Case1 and BM-Case2, due to the limited time of R18 study, specification enhancement is needed to support the following types of beam predictions:~~***

* ***~~DL Tx beam prediction with a network-side AI/ML model~~***
* ***~~DL Tx beam prediction and DL beam pair prediction with a UE-side AI/ML model~~***
  + ***~~Note: Compared to DL Tx beam prediction, no additional specification impacts for DL beam pair prediction~~***
* ***For DL beam pair prediction with a NW-side AI/ML model:***
  + ***XX companies think its spec impact is similar to DL Tx beam prediction with a NW-side AI/ML***
    - ***Xx, xxx, InterDigital - companies to add its name***
  + ***XX companies think its spec impact is higher than that of DL Tx beam prediction with a NW-side AI/ML***
    - ***Xx, xxx, Xiaomi, CMCC, CATT- companies to add its name***
* ***For DL beam pair prediction with a UE-side AI/ML model:***
  + ***XX companies think its spec impact is similar to DL Tx beam prediction with a UE-side AI/ML***
    - ***Xx, xxx, Xiaomi, CMCC, CATT, InterDigital- companies to add its name***
  + ***XX companies think its spec impact is higher than that of DL Tx beam prediction with a UE-side AI/ML***
    - ***Xx, xxx, - companies to add its name***

|  |  |
| --- | --- |
| Company | Comments |
| Mod | In order to address the concerns from some proponents of Alt.3 that cannot accept any “negative statement” (e.g., deprioritized) on DL beam pair prediction, “due to the limited time of R18 study” is added. It can imply that we don’t have enough time to complete the throughout study on DL beam pair prediction for NW-sided model, rather than saying it is infeasible or deprioritized. Hope it is acceptable. |
| LG | Intention is fine but wording seems unclear if we declare that ‘spec enhancement is needed’.. which enhancement it refers to? If the intentions is to capture this as an observation in the TR, some re-wording is suggested:  ***Proposal 3.2.1(for observation captured in the TR): For BM-Case1 and BM-Case2, ~~due to the limited time of R18 study, specification enhancement is needed to~~ most companies support the following types of beam predictions:***   * ***DL Tx beam prediction with a network-side AI/ML model*** * ***DL Tx beam prediction and DL beam pair prediction with a UE-side AI/ML model***   + ***Note: Compared to DL Tx beam prediction, no additional specification impacts may be needed for DL beam pair prediction*** |
| Xiaomi | We share similar view as LG. the intention is fine, but it is not clear. It seems try to provide some guidelines for WI. We are fine with LG’s update with a small typo fixed as below  ***Proposal 3.2.1(for observation captured in the TR): For BM-Case1 and BM-Case2, ~~due to the limited time of R18 study, specification enhancement is needed to~~ most companies support the following types of beam predictions:***   * ***DL Tx beam prediction with a network-side AI/ML model*** * ***DL Tx beam prediction and DL beam pair prediction with a UE-side AI/ML model***   ***Note: Compared to DL Tx beam prediction, additional specification impacts may be needed for DL beam pair prediction*** |
| NTT DOCOMO | Support the proposal. |
| CATT | OK |
| ZTE | Fine. The note can be deleted. It is not appropriate to conclude that there is no additional specification impacts for DL beam pair prediction compared to DL Tx beam prediction since the group have not done enough study on that. |
| Samsung | In our understanding, the intention of this proposal is to consider the spec impact of beam prediction for NW-side and UE-side model. For UE-side model, there is no additional spec impact on Rx beam determination since it is up to UE implementation.  Hence, we suggest the following modification to the proposal.  ***Proposal 3.2.1: For BM-Case1 and BM-Case2, ~~due to the limited time of R18 study,~~ consider specification enhancement ~~is needed~~ to support the following types of beam predictions:***   * ***DL Tx beam prediction with a network-side AI/ML model*** * ***DL Tx beam prediction ~~and DL beam pair prediction~~ with a UE-side AI/ML model***   ***Note: ~~Compared to DL Tx beam prediction, no additional specification impacts for DL beam pair prediction~~ For UE-side model, it is up to UE implementation to determine the Rx beam corresponding to the predicted Tx beam.*** |
| HW/HiSi | Not support. This proposal is not needed.  For both network side and UE side we already have agreements about spec impact. There is no need repeat them with a generic proposal saying that spec impact is needed.  If the intention of this proposal is here to say that no spec impact especially for beam pair prediction, then, we do not need to discuss this proposal here and instead make a conclusion:  ***Conclusion: Compared to DL Tx beam prediction, no additional specification impacts for DL beam pair prediction is studied in Rel-18.*** |
| QC | OK |
| Lenovo | For DL beam pair prediction for UE-side model, we want to clarify that “the UE is not required to report the Rx beam information”. |
| Futurewei | Support in general. Agree to add “may be needed” proposed by LG to the note; it may be too early to say there is not additional spec impacts for beam pair predictions. |
| Ericsson | Not sure what the agreement will add on top of previous arguments. We should focus on what to conclude regarding the beam pair prediction feasibility. Our view is to capture the following, to give some direction for future work.  ***Proposal 3.2.1(for observation captured in the TR)*: Conclude that beam pair prediction requires more specification impact than TX-beam prediction for NW-sided models. For UE-sided models, compared to DL Tx beam prediction, no additional specification impacts for DL beam pair prediction.** |
| Nokia/NSB | It is bit unclear what exactly the intension of the proposal. We could try to capture some observations and conclusions as suggested by few companies. |
| Spreadtrum | Agree with ZTE. Whether there are other spec impacts for beam pair prediction for UE-side model needs to be further discussed |
| Sony | For the note, we also think it can be further studied the additional spec impact, so we agree with Xiaomi's revised version. |
| MediaTek | We share the same view with ZTE, have we concluded that “Compared to DL Tx beam prediction, no additional specification impacts for DL beam pair prediction”? |
| CMCC | It seems the intention is recommend some types of beam prediction. Fine with xiaomi’s version. |
| Google | OK with HW’s proposal |
| Mod | Unfortunately, the views are quite divergent in the offline session and the inputs of the summary. A new observation is prepared to collect companies’ view, rather than what we should do, just to reflect what the fact it is. |
| Xiaomi | Input xiaomi’s view |
| CATT2 | Add CATT’s view. |
| ZTE | What does it mean by saying 'similar'？ With a similar spec impact, does it mean the spec impact of DL Tx beam prediction and beam pair prediction should be the same, or can be different? |
| HW/HiSi | Not support.  Beam pair prediction can be carried out at the UE based on implementation. We do not see any need to study spec impact for beam pair. From our understanding of the evaluations, beam pair has not shown benefits and here we should also have in mind that it has been evaluated under simplified conditions. Under more realistic conditions, the performance could be reduced further.  Considering the status of the discussion, we think it would be realistic to conclude that no consensus was achieved whether spec impact for beam pair prediction on top of DL Tx beam prediction should be introduced. |
| InterDigital | Support and added our company name with our preference. |

## AL/ML inference at UE side

In previous RAN1 meeting(s), the agreement(s)/conclusion(s) were made as below:

|  |
| --- |
| **RAN1#110bis-e**  Agreement  For BM-Case1 with a UE-side AI/ML model, study the potential specification impact of L1 signaling to report the following information of AI/ML model inference to NW   * The beam(s) that is based on the output of AI/ML model inference * FFS: Predicted L1-RSRP corresponding to the beam(s) * FFS: other information   Agreement  For BM-Case2 with a UE-side AI/ML model, study the potential specification impact of L1 signaling to report the following information of AI/ML model inference to NW   * The beam(s) of N future time instance(s) that is based on the output of AI/ML model inference   + FFS: value of N * FFS: Predicted L1-RSRP corresponding to the beam(s) * Information about the timestamp corresponding the reported beam(s)   + FFS: explicit or implicit * FFS: other information   **RAN1#112**  Agreement  For BM-Case1 and BM-Case2 with a UE-side AI/ML model, study the necessity, feasibility and the potential specification impact (if needed) of the following information reported from UE to network:   * Predicted L1-RSRP(s) corresponding to the DL Tx beam(s) or beam pair(s)   + Whether/how to differentiate predicted L1-RSRP and measured L1-RSRP * Confidence/probability information related to the output of AI/ML model inference (e.g., predicted beams)   + FFS: Definition/content of confidence/probability information * Note: At least the performance and spec impact should be considered   Agreement  For BM-Case1 and BM-Case2 with a UE-side AI/ML model, study potential specification impact of AI model inference from the following additional aspects on top of previous agreements:   * Indication of the associated Set A from network to UE, e.g., association/mapping of beams within Set A and beams within Set B if applicable * Beam indication from network for UE reception * Note: The second bullet may or may not have additional specification impact (e.g., legacy mechanism may be reused).   **RAN1#113**  Conclusion  For the study of DL beam pair prediction of BM-Case1 and BM-Case2 with a UE-side AI/ML model, RAN1 has no consensus to support the reporting of the predicted Rx beam(s) (e.g., Rx beam ID, Rx beam angle information, etc) from UE to network. |

The related proposals in tdocs are copied as below:

|  |  |
| --- | --- |
| H3C [1] | *Proposal 2: For BM-Case1 and BM-Case2 with UE-side model, further study Alt.1 DL Tx beam prediction and Alt3 Beam pair prediction.*  *Proposal 3: For BM-Case1 and BM-Case2 with a UE-side AI/ML model, study the following options for the association/mapping of beams within Set A and beams within Set B:*  *• Alt.1 Network-indicated association/mapping of beams within Set A and Set B for the UE (e.g., QCL, bitmap, formula, etc.).*  *• Alt.2 UE self-maintained association/mapping of beams within Set A and Set B (e.g., UE requests the corresponding Set B from the network).* |
| FUTUREWEI[2] | *Proposal 5: For BM-Case1 and BM-Case2 with a UE-side AI/ML model, study the necessity, and potential specification impact of AI model inference from the following additional aspects on top of previous agreements:*   * *For BM-Case1: L1 reporting of more than 4 predicted beams and the associated L1-RSRP (if applicable) in one reporting instance.* * *For BM-Case2: L1 Reporting of more than 4 predicted beams and the associated L1-RSRP (if applicable) for at least one of N time instance(s) in one reporting instance.* * *FFS: values of N (e.g., fixed or variable)* * *FFS: How to reduce the overhead*   *Note1: The performance gains should be justified by considering UCI payload overhead.* |
| Huawei[4] | *Observation 2: For the beam prediction mechanisms for BM-Case1 and BM-Case2, Alt.1 (DL Tx beam prediction) is a natural replacement of the legacy P1/P2 procedure for Tx beam sweeping, and is compatible with any pattern of the Rx beams.*  *Observation 4: For the AI/ML-based DL Tx beam prediction, non-AI/ML options can be implemented to optimize the Rx beam selection*  *• Opt.1: Fixed Rx beams is used for inference during P1/P2 and the Rx beam sweeping is performed to determine the Rx beam in P3*  *• Opt.2: A quasi-optimal DL Rx beam can be identified by sweeping the always-on SSB beams at P1 and used for Tx beam prediction at P2*  *• Opt.3: Exhaustive Rx beam sweeping is swept over multiple P1/P2 periods each of which predicts the best Tx beam for a specific Rx beam*  *Observation 5: DL Tx-Rx beam pair prediction is much more complicated to evaluate and a thorough assessment has not yet been performed. At least following issues have to be be taken into account to assess its performance:*  *• UE rotations and Rx beam blocking (when applicable)*  *• RSRP measurement errors*  *• Performance/overhead/latency*  *• Complexity*  *Proposal 7: In Rel-18 SI, not to further discuss the spec impact for DL Tx-Rx beam pair prediction on top of DL Tx beam prediction.*  *Proposal 14: For AI/ML model inference at the UE-side under BM-Case 1 and BM-Case 2, study the potential spec impact of L1 signaling to report the predicted beam IDs of more than 4 beams in one reporting instance, because*  *• It improves the beam prediction accuracy*  *• It improves the generalization performance*  *• It makes the functionality symmetric with the capabilities of a network-side model*  *Proposal 15: For AI/ML model inference at the UE-side, the motivation of introducing the report of predicted L1-RSRP(s) and confidence/probability of the AI/ML output is not clear and can be deprioritized.* |
| Spreadtrum[5] | *Proposal 1: For sub use cases BM-Case1 and BM-Case2,*  *• If Set A and Set B are different, the association between set B and set A should be configured by gNB.*  *• If Set B is a subset of Set A, the association between set B and set A can be determined by beam index.*  *Proposal 2: For sub use cases BM-Case1 and BM-Case2, support both Alt 1 and Alt 3.*  *Proposal 11: For BM-Case1 and BM-Case2 with a UE-side AI/ML model, study the enhancement for beam report without RSRP.* |
| IDC[6] | *Observation 3: Reporting information of multiple future time instances in one report can reduce latency.*  *Proposal 3: For BM-Case 2 with a UE-side AIML model, support reporting measurements of multiple future time instances in one report.*  *Proposal 4: To indicate Set A beams not in Set B, consider an indication method based on Set B information (e.g., Set B beam IDs).*  *Observation 14: For BM-Case 1 with a UE-side AI/ML model, the current beam reporting with CRIs/SSBRIs and corresponding L1-RSRP values is enough to indicate the best beam(s) from AI/ML model inference.*  *Observation 15: For BM-Case 2 with a UE-side AI/ML model, information about the time stamp for the reported CRIs/SSBRIs can be further considered.*  *Proposal 23: For a UE-side AI/ML model, consider information about the time stamp for potential specification impact.*  *Observation 16: Reporting confidence/probability information per predicted beam can cause unnecessary reporting overhead.*  *Proposal 24: Consider reporting confidence/probability information related to the output of AI/ML model per an output instance.*  *Proposal 25: Consider using legacy procedures to indicate the mapping between Set A and Set B to the UE.* |
| Vivo[7] | *Proposal 34: For BM-Case1 and BM-Case2 with a UE-side AI/ML model, study potential specification impact on resource configuration for AI/ML model inference:*  *• Specific beam pair resource configuration for Set B/Set C*  *• Enhanced P3+P2 resource configuration that Rx beam assumption of P2 resource measurement is the best Rx beam searched from P3 procedure for Set B/Set C*  *Proposal 35: For BM-Case1 and BM-Case2 with a UE-side AI/ML model, study potential specification impact on beam resource request for AI/ML model inference:*  *• Renew beam pattern request w or w/o beam pattern suggestion from UE to NW*  *• Minimum resource number request from UE to NW*  *- Minimum number of requested beams*  *- Minimum number of requested repetitions*  *Proposal 36: For BM-Case1 and BM-Case2 with a UE-side AI/ML model, study potential specification impact on assistance information for AI/ML model inference:*  *• Proprietary processed Tx beam information as assistance information from NW to UE*  *Proposal 37: For BM-Case1 and BM-Case2 with a UE-side AI/ML model, study potential specification impact on beam report for AI/ML model inference:*  *• Predicted L1-RSRP report and study how to report predicted beam indicator*  *• Fallback beam report to indicate invalid measured results for AI/ML based beam prediction*  *• Study how to further reduce report overhead of time domain beam prediction for predicted results of multiple occasions.* |
| ZTE[8] | *Proposal 8: For BM-Case1 and BM-Case2 with a UE-side AI/ML model, if Set B and Set A are different, support to indicate the association of Set A and Set B from NW to UE at least for model inference.*  *Proposal 9: For BM-Case1 and BM-Case2 with a UE-side AI/ML model, if Set B is a subset of Set A, support to indicate the mapping of Set A and Set B from NW to UE in the form of pattern ID or beam identifier at least for model inference.*  *Proposal 11: For BM-Case1 and BM-Case2 with a UE-side AI/ML model, study enhanced UE reporting to allow the reporting of unmeasured beams in set A.*  *Proposal 12: For BM-Case1 and BM-Case2 with a UE-side AI/ML model, support the reporting of both predicted RSRP and confidence information for beam selection at the NW side.*  *Proposal 13: For BM-Case1 and BM-Case2 with a UE-side AI/ML model, if both the predicted RSRP and measured RSRP to the same beam are available at the UE side, the measured RSRP should be reported due to its higher reliability.*  *Observation 5: NW can differentiate the predicted RSRP and measured RSRP based on whether or not the reported beam is from Set B.*  *Observation 6: Since the predicted RSRP and measured RSRP of different beams may be reported in one reporting instance, the beam with lower measured RSRP may be associated with higher predicted RSRP in case of prediction error.*  *Proposal 14: For BM-Case1 and BM-Case2 with a UE-side AI/ML model, study enhanced reporting mechanism to support the reporting of the predicted RSRP or measured RSRP for different beams in the same reporting instance.*  *Proposal 15: For BM-Case1 and BM-Case2 with a UE-side AI/ML model, L1 reporting of the beam indicators should be based on the AI/ML model output.*  *Proposal 16: For BM-Case1 and BM-Case2 with a UE-side AI/ML model, if the associated RSRP is to be reported, the necessity of reporting more than 4 predicted beams should be evaluated in agenda 9.2.3.1 first to see if enough performance gains can be achieved.*  *Proposal 17: For BM-Case1 and BM-Case2 with a UE-side AI/ML model, if the confidence/probability information is to be reported, the necessity of reporting more than 4 predicted beams highly depends on how to determine the reported beams.*  *Proposal 18: For BM-Case1 and BM-Case2 with a UE-side AI/ML model, study reporting of prediction results of multiple future time instance(s) in one reporting instance.*  *Proposal 19: For BM-Case1 and BM-Case2 with a UE-side AI/ML model, whether the number of future time instances to be predicted can be variable should be evaluated in agenda 9.2.3.1 first to see if enough performance gains can be obtained.*  *Observation 9: Since the data collection of beam pair prediction incorporates Tx beam sweeping and Rx beam sweeping simultaneously, it may take a long time for one round of data collection for model inference.*  *Observation 10: As the UE Rx beam is up to implementation, the P1 beam sweeping procedure is still conceptual and there is no explicit signaling/configuration for P1 in current specification.*  *Proposal 21: Study enhanced resource configuration for P1 beam sweeping procedure to facilitate a timely data collection for model inference of UE-side beam pair prediction.*  *Observation 11: The predicted Top-K beam pairs may include two or more Rx beams that is associated with the same Tx beam.*  *Proposal 22: To differentiate the multiple beam pairs specific to the same Tx beam in the UE reporting, the Tx beam associated with different Rx beams can be reported repeatedly.*  *Proposal 23: Considering UE has better knowledge on the confidence level of the predicted top-1 or top-K beams, the additional RS resource for the second stage beam sweeping can be requested by UE.*  *Observation 12: Without Rx beam information reporting, the QCL type D relationship associated with each resource for top-K beam sweeping may not be available and thus the RS resource overhead for the second stage beam sweeping may be significantly increased.*  *Proposal 30: If the currently activated model is not able to accurately predict the optimal beam information, the UE reporting of beam ID or beam quality information should be based on measurement results of Set B instead of prediction results of Set A.* |
| Intel[9] | *Proposal 1: Beam Pair prediction (Alt-3) should be supported, at least for BM-Case 1 since it can provide large latency and measurement gains for joint P2/P3 procedure*  *Observation 1: For beam pair prediction at UE side using DL measurements, if gNB provides indexes ordered in terms of angular coverage of beams, there may be no need to divulge proprietary information and UE can use the beam indexes to construct set B for input to ML model.*  *Proposal 8: If the predicted RSRP is directly obtained from the output of the AI/ML model, such metrics should not be reported back to the network or should be reported with an indication that the report is a prediction from UE side model and not an actual measurement.*  *Proposal 9: If the predicted RSRP is defined as the actual measured RSRP on the best beam predicted by the model, such metric can be reported back to the network assuming that the overhead due to the additional measurement, if needed, is accounted for in the KPI calculation.*  *Proposal 10: For NW-side AI/ML model, support beam/L1-RSRP reporting over MAC-CE for reports spanning multiple time instances.* |
| Ericsson[11] | *Proposal 9 For BM-Case1 and BM-Case2 with a UE-side AI/ML model, study potential specification impact of beam indication overhead reduction from the following additional aspect*  *a. Enhanced CSI resource/report configuration, how to adapt the TCI switch time offsets or configure several TCIs in one configuration.*  *Proposal 16 Conclude that the specification impact for DL beam pair prediction at UE sided model inference is same as for TX DL beam prediction* |
| Google[12] | *Proposal 5: Do not support spec impact for L1-RSRP prediction.* |
| LG[13] | *Proposal #5: Support predicted L1-RSRP report together with beam(s). For BM-Case2, information on time-variation of L1-RSRP can also be included in the report for helping intra-/extra-polation at NW side.*  *Proposal #6: For predicted L1-RSRP report, confidence/probability information may be helpful for NW to decide whether/how to use the reported L1-RSRP. Further study whether the information is per model/functionality, per report or per report parameter.*  *Proposal #7: For BM-Case2 with UE-sided models, following beam reporting enhancements can be considered*  *- Report of beam(s) for each future time instance or beam(s) for a time duration, i.e. from the first time instance to the last time instance.*  *- Report of beam(s) for current time instance for fallback operation*  *- Report of timestamps by UE or NW to indicate timestamps* |
| CATT[14] | *Proposal 12: For BM-Case1 and BM-Case2 with a UE-side AI/ML model, regarding how to indicate the association/mapping of Set A and Set B:*  *• Study indicating the resources of Set A and Set B associated with CSI reporting*  *• Study both explicit indication and implicit indication methods.* |
| NEC[15] | *Proposal 7: Regarding model inference for BM-Case1/2 with UE-side AI/ML model, study to report the number of predicted beams to report in beam report.*  *Proposal 8: Regarding model inference for BM-Case1/2 with UE-side AI/ML model, study to report the number of predicted beams (for finding the actual best beam) to NW if it is determined at UE side.*  *Proposal 9: Regarding model inference for BM-Case2 with UE-side AI/ML model, study to report the number of future time instances to NW if it is determined at UE side.* |
| Fujitsu[16] | *Proposal 7: For DL beam (pair) prediction with a UE-side AI/ML model, study the potential specification impacts of model inference on*  * The request to NW about the required RSs of Set B*  * The association/mapping between Set B and Set A*  * FFS on the method, e.g., the bitmap or QCL to set the association/mapping between Set B and Set A.* |
| CMCC[17] | *Proposal 1: Regarding beam prediction type, Tx beam prediction at NW side is prioritized, Tx beam/beam pair prediction at UE side can be considered, beam pair prediction at NW side is not considered.*  *Proposal 10: For BM-Case1 with a UE-side AI/ML model, study the following L1 beam reporting enhancement for AI/ML model inference*  *• How to indicate association/mapping of beams within Set A and beams within Set B from NW to UE*  *· Opt1: QCL information between beams within Set A and beams within Set B*  *· Opt2: Identifier of info representing the association/mapping of Set A and Set B*  *e.g. start and ending indicator*  *· Opt3: The bitmap or pre-defined rule that beams of Set B is subset of beams of Set A*  *Proposal 11: For BM-Case1 with a UE-side AI/ML model, study the following L1 beam reporting enhancement for AI/ML model inference*  *• L1 reporting of more than 4 predicted beams and the associated L1-RSRP (if applicable) in one reporting instance*  *• additional spec impact of beam pair prediction compared to DL Tx beam prediction*  *Proposal 12: For BM-Case1 with a UE-side AI/ML model, whether the predicted L1-RSRP is reported can be configured by the gNB, whether/how to differentiate measured L1-RSRP and predicted L1-RSRP needs further discussion.*  *Proposal 17: For BM-Case2 with a UE-side AI/ML model, study the necessity, benefit(s), and potential specification impact of AI model inference from the following additional aspects on top of previous agreements:*  *• For BM-Case2: L1 Reporting of more than 4 predicted beams and the associated L1-RSRP (if applicable) for at least one of N time instance(s) in one reporting instance*  *o FFS: values of N (e.g., fixed or variable)*  *• FFS: How to reduce the overhead* |
| Panasonic[18] | *Proposal 2: CSI reporting framework can be considered as starting point for UE to report beam prediction to NW in case of UE-side inference.*  *Proposal 3: Prediction related metrics can be introduced in the CSI report configuration as the report quantities. FFS the following prediction related metrics:*  *- Predicted beam ID (or RS ID, or TCI State ID)*  *- Predicted beam quality, such as predicted L1-RSRP, L1-SINR*  *- Predicted beam application time (when to start/stop applying the predicted beam)*  *- Confidence/probability information*  *Proposal 4: To distinguish between prediction and measurement results, following options can be considered:*  *- NW indicates explicitly prediction-related metrics in the CSI report configuration*  *- NW configures different resource set in CSI report configuration for prediction and measurement, respectively.* |
| Nokia[19] | *Proposal 13. For UE-sided BM-Case1 a with a UE-side AI/ML model, consider the potential specification impact of L1 signaling to report predicted L1-RSRP to the NW.*  *Observation 5. To distinguish predicted L1-RSRP from measured L1-RSRP when the UE-sided model is employed,*  *• If a reported beam belongs to Set B, NW knows it is a measured L1-RSRP, otherwise, NW knows it is a predicted L1-RSRP.*  *Proposal 14. RAN1 to consider reporting confidence/probability information related to the output of AI/ML model inference (e.g., predicted beams).*  *Proposal 15. RAN1 to consider overhead reduction on predicted beams by means of configuring variable value of N beams for predicted beams.* |
| Xiaomi[12] | *Proposal 8: Consider one absolute L1-RSRP for each time instance or one absolute L1-RSRP for all time instance in one beam report including beam reports of more than one time instance for BM-case 2.*  *Proposal 9: Consider UE to report the number/ periodicity of the time instance in beam report for BM-case 2.* |
| DCM[22] | *Proposal 11: Study if the legacy CSI framework, such as CSI processing unit framework and CSI computation time mechanism, can be reused for predicted beam reporting.* |
| OPPO[23] | *Observation 6: For Tx beam prediction (Alt.1), the corresponding Rx beam could either be determined by UE as specific Rx beam or by existing Rx beam sweeping procedure.*  *Proposal 6: For Tx beam prediction (Alt.1) with UE-side model, legacy beam reporting and indication mechanism could be reused.*  *Observation 7: For beam pair prediction, legacy beam reporting and indication may cause Rx beam confusion (e.g. indicated Tx beam corresponding to different Rx beams).*  *Observation 8: For beam pair prediction, the conclusion of no reporting of Rx beam info. will impact beam indication mechanism, e.g. via legacy TCI framework to convey Tx beam info. only.*  *Observation 9: The reported predicted L1-RSRP and measured L1-RSRP can be differentiated via separate NW configuration.*  *Proposal 7: Predicted L1-RSRP by UE-side model should be reported to NW along with predicted Top-K Tx beam(s) or beam pair(s).*  *Proposal 8: Confidence/probability of UE-side model output could be quantized and reported to NW.*  *Observation 10: From signaling aspects, it seems flexible to configure both Set A and Set B via higher layer signaling.*  *Proposal 9: For BM-Case2 with UE-side model, UE reports the predicted beam (pair) for N future time instance(s) by single reporting instance.*  *Proposal 10: For BM-Case2 with UE-side model, the timestamp of N future time instance(s) should be implicitly reported to NW.*  *Proposal 11: For BM-Case2, NW indicates multiple beam indications for future N time instances.* |
| Samsung[24] | *Proposal 14. For BM-Case1 with a UE-side AI/ML model, for model inference, at least the following with potential specification impact is identified:*  * gNB provides the configurations associated with Set A and/or Set B*  * L1 beam report with predicted beams (or beam pairs)*  * L1 beam report with predicted L1-RSRP(s) corresponding to the predicted beams (or beam pairs), if applicable*  *Proposal 15: For BM-Case1 with a UE-side AI/ML model, the following potential specification impacts can be identified:*  * Indication of beams in Set A not in Set B using the indicated information of Set A*  *Proposal 28. For BM-Case2 with a UE-side AI/ML model, for model inference, at least the following with potential specification impact is identified:*  * Reporting information about the predictions of N future time instances in one reporting instance for BM-Case2*  *Proposal 29: For BM-Case2 with a UE-side AI/ML model, at least the following with potential specification impact is identified for time domain predictive beam indication:*  * A single beam indication for multiple future time instances* |
| Rakuten[25] | *Observation 1: In a UE-sided model of AI-ML based beam prediction, Set A of beams cannot be always assumed to have been transmitted earlier to prediction.*  *Observation 2: In a UE-sided model of AI-ML based beam prediction, the AI-ML model output parameters (beam Id, L1 RSRP) provided by UE are not sufficient to associate and identify all the predicted beams at gNB.*  *Proposal 1: RAN1 agrees the use case where, in a UE-sided model, beams of Set A cannot be always assumed to be transmitted before the prediction.*  *Proposal 2: RAN1 agrees to define Set A to Set B association differently for transmitted and un-transmitted beams before prediction.*  *Proposal 3: RAN1 discusses and agrees a solution to associate and identify un-transmitted beams predicted by UE in a UE-sided model.*  *Observation 3: In a UE-sided model, the AI-ML model output parameters (beam Id, L1 RSRP) provided by UE are not sufficiently granular enough to schedule UE-specific data beams.*  *Proposal 4: Tx and/or Rx Beam angle(s) of Set A beams are defined and predicted with more granularity to enable beamforming of precise and accurate data beams.* |
| ETRI[26] | *Proposal 3. For BM-Case1 and BM-Case2 with a UE-side AI/ML model, an approach using AI/ML model ID for inferring the relationship between Set A and Set B can be considered.* |
| Lenovo[27] | *Proposal 1: Study the potential specification impact for the UE to report or obtain the RSRPs for a same Tx beam with different Rx beams for the AI/ML Model input for NW-side or UE-side AI/ML inference.*  *Proposal 12: Rel-17 CSI reporting framework can be reused for UE-side beam prediction by configuring measurement beam Set B as the channel measurement resource, but the reported beam is selected from another prediction beam Set A.* |
| CAICT[28] | *Proposal 3: For AI/ML model inference at the UE-side under BM-Case 1 and BM-Case 2, the reporting number of beam information (i.e. beam IDs/RSRPs) should be configurable and adaptive to reduce reporting overhead.* |
| KT[29] | *Proposal 4. For the model training/inference/monitoring of the UE-side AI/ML model:*  *• Study how to identify the beams within Set A which is not measured.*  *• Study how to indicate the association/mapping of beams within Set A and beams within Set B.* |
| QC[30] | *Proposal 4: For BM-Case2 with a UE-side AI/ML model, study the potential specification impact of L1 signalling to report the following information of AI/ML model inference to NW:*  *• Predicted beam blockage/failure* |
| MTK[31] | *Proposal 22: For BM-Case2, instead of configuring periodic beam report, NW can configure separate reports for each BM-Case2 model inference instance.*  *Proposal 24: For BM-Case1 and BM-Case2 with a UE-side AI/ML model, study spec impact to facilitate UE to report various number of Top-K beams in one beam report (K ≤ nrofReportedRS) as AI/ML model output.* |
|  |  |

###### Mod’s assessment

**Indication of association/mapping of Set A and Set B.**

At least 8 companies (e.g., H3C, Spreadtrum, IDC, ZTE, CATT, Fujitsu, CMCC, ETRI) are discussing the indication of association/mapping of Set A and Set B.

* **Mod’s assessment:** Some of the proposals are quite generic and lack details. The other proposals are quite divergent. Similar situation as before. It is very difficult to formulate a proposal to capture the proposed solutions. Moreover, this discussion is related to some other issues. Thus, moderator feels it is premature to discuss the detailed solutions at current stage.
* **Mod’s suggestion:** No discussion in this meeting. The detailed solutions can be discussed in R19 WI (if any).
* **Related proposals in tdocs**
  + H3C: Proposal 3
  + Spreadtrum: Proposal 1
  + IDC: Proposal 25
  + ZTE: Proposal 8, 9
  + CATT: Proposal 12
  + Fujitsu: Proposal 7
  + CMCC: Proposal 10
  + ETRI: Proposal 3
  + LGE: Proposal 1

**Reporting of more than 4 predicted beams**

At least 3 companies (e.g., Futurewei, Huawei, CMCC) suggest to the support of more than 4 predicted beams and associated L1-RSRP (if applicable) in one reporting. In contrast, one company (ZTE) thinks it is unnecessary.

* **Mod’s assessment:** The corresponding proposal was discussed in the last two meetings. A number of companies thought it is unnecessary. Meanwhile, some proponents of this proposal agreed it can be discussed in WI.
* **Mod’s suggestion:** No discussion in this meeting. It can be further discussed in R19 WI (if any)
* **Related proposals in tdocs**
  + Futurewei: Proposal 5
  + Huawei: Proposal 14
  + CMCC: Proposal 11, 17
  + ZTE: Proposal 16

**Other issues**

There are also some proposals for other issues, e.g., reporting of predicted beam failure, predicted beam applicable time, how to select the top-K beams, overhead reduction and so on.

* **Mod’s suggestion:** Since there are usually one or two companies discussing each issue, we can wait for more concrete inputs/proposals from other companies for each of these issues.

# Spec impact of Model monitoring

## General aspects

In previous RAN1 meeting(s), the agreement(s)/conclusion(s) were made as below:

|  |
| --- |
| **RAN1#110**  Agreement  Regarding the model monitoring for BM-Case1 and BM-Case2, to investigate specification impacts from the following aspects   * Performance metric(s) * Benchmark/reference for the performance comparison * Signaling/configuration/measurement/report for model monitoring, e.g., signaling aspects related to assistance information (if supported), Reference signals * Other aspect(s) is not precluded   **RAN1#110bis-e**  Agreement  Study AI/ML model monitoring for at least the following purposes: model activation, deactivation, selection, switching, fallback, and update (including re-training).  FFS: Model selection refers to the selection of an AI/ML model among models for the same functionality. (Exact terminology to be discussed/defined)  Agreement (AI 9.2.1)  Study at least the following metrics/methods for AI/ML model monitoring in lifecycle management per use case:   * Monitoring based on inference accuracy, including metrics related to intermediate KPIs * Monitoring based on system performance, including metrics related to system ignalling KPIs * Other monitoring solutions, at least following 2 options.   + Monitoring based on data distribution     - Input-based: e.g., Monitoring the validity of the AI/ML input, e.g., out-of-distribution detection, drift detection of input data, or ~~something simple like checking~~ SNR, delay spread, etc.     - Output-based: e.g., drift detection of output data   + Monitoring based on applicable condition   Note: Model monitoring metric calculation may be done at NW or UE  Agreement (AI 9.2.1)  Study performance monitoring approaches, considering the following model monitoring KPIs as general guidance   * Accuracy and relevance (i.e., how well does the given monitoring metric/methods reflect the model and system performance) * Overhead (e.g., ignalling overhead associated with model monitoring) * Complexity (e.g., computation and memory cost for model monitoring) * Latency (i.e., timeliness of monitoring result, from model failure to action, given the purpose of model monitoring) * FFS: Power consumption * Other KPIs are not precluded.   Note: Relevant KPIs may vary across different model monitoring approaches.  FFS: Discussion of KPIs for other LCM procedures  **RAN1#112**  Agreement  Regarding the performance metric(s) of AI/ML model monitoring for BM-Case1 and BM-Case2, study the following alternatives (including feasibility/necessity) with potential down-selection:   * Alt.1: Beam prediction accuracy related KPIs, e.g., Top-K/1 beam prediction accuracy * Alt.2: Link quality related KPIs, e.g., throughput, L1-RSRP, L1-SINR, hypothetical BLER * Alt.3: Performance metric based on input/output data distribution of AI/ML * Alt.4: The L1-RSRP difference evaluated by comparing measured RSRP and predicted RSRP * Other alternatives are not precluded * Note: At least the performance and spec impact should be considered   **RAN1#112bis-e**  Agreement  For AI/ML performance monitoring for BM-Case1 and BM-Case2, study potential specification impact of at least the following alternatives as the benchmark/reference (if applicable) for performance comparison:   * Alt.1: The best beam(s) obtained by measuring beams of a set indicated by gNB (e.g., Beams from Set A)   + FFS: gNB configures one or multiple sets for one or multiple benchmarks/references * Alt.4: Measurements of the predicted best beam(s) corresponding to model output (e.g., Comparison between actual L1-RSRP and predicted RSRP of predicted Top-1/K Beams) * FFS:   + Alt.3: The beam corresponding to some or all the indicated/activated TCI state(s) * Other alternative is not precluded. |

The related proposals in tdocs are copied as below:

|  |  |
| --- | --- |
| NVIDIA[3] | *Proposal 7: For AI/ML based beam prediction in spatial/time domain, study potential specification impact related to assistance signalling and procedure for model performance monitoring and model update/tuning.* |
| Huawei[4] | *Proposal 12: For the performance metrics of monitoring, the discussion on spec impact for input or output data based monitoring (Alt.3) and L1-RSRP difference evaluated by comparing measured and predicted RSRP (Alt.4) can be deprioritized.* |
| Spreadtrum[5] | *Proposal 12: For BM-Case1 and BM-Case2, the RSRP difference evaluated by comparing actual RSRP and predicted RSRP can be used as a performance metric.* |
| IDC[6] | *Proposal 17: For AI/ML monitoring, consider a common mechanism for multiple purposes, procedures for identifying need of AI/ML model recovery, UE request/gNB trigger and AI/ML model recovery.*  *Observation 12: Supporting only one KPI for AI/ML model monitoring may not work in some potential scenarios.*  *• E.g., best predicted beam based monitoring (Alt.1) may not work if qualities of adjacent beams are similar.*  *Observation 13: Supporting multiple KPIs can be beneficial as each KPI has different functionalities.*  *Proposal 18: Support both Alt.1 and Alt.4 for AI/ML model monitoring and consider applying different KPIs considering implementation scenarios.*  *• Alt.1: The best beam(s) obtained by measuring beams of a set indicated by gNB (e.g., Beams from Set A).*  *• Alt.4: Measurements of the predicted best beam(s) corresponding to model output (e.g., Comparison between actual L1-RSRP and predicted RSRP of predicted Top-1/K Beams).* |
| Vivo[7] | *Proposal 41: Regarding the performance metric(s) of AI/ML model monitoring for BM-Case1 and BM-Case2, support Alt.1 and Alt.4,*  * Alt.1: Beam prediction accuracy related KPIs, e.g., Top-K/1 beam prediction accuracy*  * Alt.4: The L1-RSRP difference evaluated by comparing measured RSRP and predicted RSRP*  *Proposal 42: Regarding the performance metric(s) of AI/ML model monitoring for BM-Case1 and BM-Case2, deprioritized Alt.3, i.e. performance metric based on input/output data distribution of AI/ML*  *Proposal 43: Support Alt.1 (the best beam(s) obtained by measuring beams of a set indicated by gNB) and Alt 4 (Measurements of the predicted best beam(s) corresponding to model output), as the benchmark/reference for performance comparison for AI/ML model monitoring for BM-Case1 and BM-Case2. Deprioritize other alternatives.* |
| ZTE[8] | *Proposal 24: All alternatives of performance metrics for AI/ML model monitoring should be evaluated in agenda 9.2.3.1 before further down-selection.*  *Proposal 25: Prioritize beam prediction accuracy related KPIs (i.e., Alt.1 and Alt.4) as the performance metric for AI/ML model monitoring since it has been evaluated in agenda 9.2.3.1 and could well reflect the performance of the AI/ML model.*  *Proposal 26: Study performance monitoring mechanism on the basis of beam failure recovery mechanism in the current specification.*  *Observation 13: Considering ping pong effect in wireless communication, the performance of the AI/ML model may change with time and inference errors occur even for a valid AI/ML model.*  *Proposal 27: Model/functionality failure detection should be based on failures on several consecutive times, or the error probability of the model/functionality failures exceeds a certain threshold during a predefined monitoring window.*  *Proposal 28: Study candidate model/functionality identification mechanism with low resource consumption after detecting the currently activated model/functionality is no longer suitable.* |
| Sony[10] | *Proposal 7 : Support Alt.2 and Alt.4 as the performance metric(s) of AI/ML model monitoring for BM-Case1 and BM-Case2.*  *Proposal 8 : Compare the RSRP of the actual best beam achieved by the traditional beam selection mechanism and the predicted best beam to determine if the functionality/model is still suitable.* |
| Ericsson[11] | *Proposal 11 Consider Table 1 for summarizing the pros/cons on different performance metrics*  **Table 1** **Summary of different performance metrics-based methods for AI/ML BM model monitoring**   |  |  |  | | --- | --- | --- | | **Performance metric** | **Benefits** | **Challenges** | | **Inference Accuracy**  (Intermediate KPIs)  (Alt.1, Alt. 4) | -Metric reflects the model performance very well  -Expected to provide reliable model failure detection | -Signalling overhead for collecting ground truth data at UE/NW (RS transmission and/or UE reporting)  -Frequent monitoring degrades the usability of the model. | | **System/Link performance metric(s)**  (Alt.2) | -Metric reflects the system performance  -Low complexity and signaling overhead | -Challenging to identify that the degradation is due to an inaccurate model  (inaccurate model monitoring) | | **Data distribution**  (Alt.3) | -No additional signaling overhead for obtaining input/output data  -Shorter latency for obtaining data samples for model monitoring  -Frequent monitoring possible | May not reflect model performance as well as Alt.1  May not reflect system performance as well as Alt. 2  To achieve reliable model failure detection, many samples may be required to calculate statistical metrics. | |
| Google[12] | *Proposal 7: For spatial domain beam prediction, the beam quality for current beam from an indicated TCI can be used for performance validation, and if none of the predicted beam(s) can provide better beam quality than current beam, the predicted beam(s) are assumed to fall to pass the performance validation.*  *Proposal 13: For time-domain beam prediction, the beam quality for current beam from an indicated TCI can be used for performance validation, and if none of the predicted beam(s) can provide better beam quality than current beam, the predicted beam(s) are assumed to fall to pass the performance validation.*  *Proposal 14: Study UE feedback before the beam action time for performance validation for predicted beam in addition to the ACK/NACK for the TCI update signaling.* |
| CATT[14] | *Proposal 14: For AI/ML performance monitoring for BM-Case1 and BM-Case2, study when to transmit the RS corresponding to the predicted Top-1/K beams if the L1-RSRP difference (Alt.4) is used as performance monitoring metric.*  *Proposal 15: For AI/ML performance monitoring for BM-Case1 and BM-Case2, study potential specification impact of the following alternatives as the benchmark/reference (if applicable) for performance comparison*  *• The best beam(s) obtained by measuring beams of multiple sets that nearly constitutes set A*  *• AI/ML solution subject to an inactive model, to make the decision of switching/selection based on the performance comparison with the AI/ML solution being monitored.* |
| CMCC[17] | *Proposal 13: For model inference of BM-Case1, beam prediction accuracy related KPI can be used as the metric of model performance monitoring.* |
| DCM[22] | *Observation 1: Beam prediction accuracy related KPI requires the large measurement overhead for searching out the actual top 1/K beam(s).*  *Observation 2: Link quality related KPI is affected by various factors in addition to model performance.*  *Proposal 4: Discuss the feasibility of the performance monitoring based on the input/output data distribution in the beam prediction, before the specification impact discussion related to it.*  *Observation 3: The required measurement overhead for performance monitoring is relatively low for the monitoring based on the predicted L1-RSRP difference.*  *Proposal 5: Study performance metric calculation per prediction time offset for monitoring temporal beam prediction.*  *Observation 6: Potential specification impacts from the performance monitoring via the measurement of the actual best beam and the predicted best beam can be summarized in Table 2.*  Table 2. Potential specification impact for Alt.1 and Alt.4 performance monitoring.   |  |  |  | | --- | --- | --- | |  | Alt.1 Beam prediction accuracy related KPI (e.g., top1/K beam prediction accuracy, L1-RSRP difference) | Alt.4 Predicted L1-RSRP difference of top1/K beam(s) | | Data collection | ・Measurements of all Set A beams  ・RS Configuration to enable the above measurement | ・Measurements of top-1/K predicted Set A beams  ・RS Configuration to enable the above measurement | | Reporting | ・Report the measurement of top-1/K measured Set A beams for NW side performance monitoring  ・Report top-1/K beam prediction accuracy/L1-RSRP difference/event occurrence for hybrid performance monitoring  ・Report upcoming operation request based on the monitored performance for UE side performance monitoring | ・Report the measurement of top-1/K predicted Set A beams for NW side performance monitoring  ・Report predicted L1-RSRP difference/event occurrence for hybrid performance monitoring  ・Report upcoming operation request based on the monitored performance for UE side performance monitoring | |
| OPPO[23] | *Proposal 18: For performance metric of AI/ML model monitoring, further study (Alt.1) the LCM mechanism based on beam prediction accuracy, e.g by comparing predicted beam and measured beam(s) (beam pair(s)) pre-configured in a set.* |
| Samsung[24] | *Proposal 17. For the performance metric(s) of AI/ML model monitoring, the necessity and feasibility of Alt-3 (i.e., performance metric based on input/output data distribution of AI/ML) should be further studied.* |
| Lenovo[27] | *Proposal 10: Select Alt 1 and Alt 4 as the performance metric(s) of AI/ML model monitoring.*  * Alt.1: Beam prediction accuracy related KPIs, e.g., Top-K/1 beam prediction accuracy.*  * Alt.4: The L1-RSRP difference evaluated by comparing measured RSRP and predicted RSRP.* |
| IITM[32] | *Proposal 1: Study the conditions and UE performance metrics based on which model LCM is triggered.*  *Proposal 2: To study metrics that can be measured at the UE for Performance monitoring* |
|  |  |

###### Mod’s assessment

**Down-selection of performance metrics**

At least 11 companies (e.g. vivo, ZTE, CMCC, OPPO, Lenovo, Sony, Huawei, DCM, Samsung, Spreadtrum, CATT) are discussing the down-selection/prioritization of the alternatives for performance metrics. Meanwhile, 1 company (Ericsson) analyzes and summarizes the advantages and disadvantages of each alternatives.

* **Mod’s assessment:** According to previous discussions, Alt.3 has the least supporting companies but the most opposing companies. Based on the tdocs, the situation is not changed. It would be good if the group can make some consensus on the down-selection so that we can have a short list of alternatives and make our future work more focused.
* **Mod’s suggestion:** The same proposal of the last meeting is suggested for further discussion.
* **Related proposals in tdocs**
  + Vivo: Proposal 41, 42
  + ZTE: Proposal 25
  + CMCC: Proposal 13
  + OPPO: Proposal 18
  + Lenovo: Proposal 10
  + Sony: Proposal 7
  + Huawei: Proposal 12
  + DCM: Proposal 4
  + Samsung: Proposal 17
  + Spreadtrum: Proposal 12
  + CATT: Proposal 14
  + Ericsson: Proposal 11

**Other issues**

There are also some proposals for other issues (e.g., model/functionality failure detection, down-selection of the alternative for benchmark/reference, and so on)

* **Mod’s assessment:** Most of these proposals are suggested by limited companies. Thus, from moderator’s perspective, it is difficult to formulate some stable proposal(s).
* **Mod’s suggestion:** No discussion until there are more supporting companies.

###### Proposal 4.1.1 (FL4)

***~~Proposal 4.1.1: Regarding the performance metric(s) of AI/ML model monitoring for BM-Case1 and BM-Case2, deprioritize Alt.3:~~******~~Performance metric based on input/output data distribution of AI/ML~~***

***Proposal 4.1.1 (Additional rows):***

***Observation:***

***Regarding the performance metric(s) of AI/ML model monitoring for BM-Case1 and BM-Case2, the following table is identified***

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Alt.1: Beam prediction accuracy related KPIs, e.g., Top-K/1 beam prediction accuracy | Alt.2: Link quality related KPIs, e.g., throughput, L1-RSRP, L1-SINR, hypothetical BLER | Alt.3: Performance metric based on input/output data distribution of AI/ML | Alt.4: The L1-RSRP difference evaluated by comparing measured RSRP and predicted RSRP |
| Signaling/RS overhead | Larger than other alternatives | Less than Alt.1/4 | ~~Less than other alternatives~~  Less than Alt.1/4 | Less than Alt.1 |
| Specification impact | Configuration/measurement/reporting for KPI calculation  Reporting of calculated KPI (if applicable) | Measurement/reporting (e.g., L1-RSRP) (if applicable) | Configuration/measurement for KPI calculation  Reporting of calculated KPI (if applicable) | Configuration/Measurement/reporting of Top-K predicted beam  Reporting of calculated KPI (if applicable) |

|  |  |
| --- | --- |
| Company | Comments |
| LG | What it means by ‘deprioritize Alt3’? Is it for deprioritizing the alt for the rest of SI?  In general, due to the remaining time for this SI, it would be better to capture it as a form of observation if really needed, e.g. ‘some companies concerned on Alt3 due to …. ‘. |
| Xiaomi | Fine with the proposal |
| NTT DOCOMO | Support the proposal. |
| CATT | OK |
| ZTE | Fine. |
| Samsung | Disagree with the proposal. In our view, Alt-3 is an attractive performance metric that can save the signaling overhead of performance monitoring. More details can be found in our discussion paper for 9.3.2.1 (R1-2307670).  For high quality TR, it is important to discuss the pros/cons for each alternative rather than de-prioritize the one single alternative at this stage.  For the discussion of pros and cons of each alternative, please see our proposal as below.  **For BM-Case1 with a network-side AI/ML model, for the alternatives for performance metric(s) of model monitoring (agreed in RAN1#112), the following observation is made:**   * **Alt.1: Beam prediction accuracy related KPIs, e.g., Top-K/1 beam prediction accuracy**   + **This can be done by comparing actual beam measurement and beam prediction from Set A**   + **This may require additional support of L1 beam report of the measurement results of more than 4 beams in one reporting instance**   + **This may require additional support of L1 beam report of the measurement results with beam index only (e.g., without the corresponding L1-RSRP)**   + **This may not reflect system performance since it is an intermediate KPI** * **Alt.2: Link quality related KPIs, e.g., throughput, L1-RSRP, L1-SINR, hypothetical BLER**   + **This is an indirect metric for model monitoring, which makes it prone to the potential impact of other factors (e.g., bad linkage quality)**   + **This reflects system performance;** **however, it is hard to differentiate the poor performance is due to wrong selection of the beams(pairs) or due to other factors (e.g., bad linkage quality)** * **Alt.3: Performance metric based on input/output data distribution of AI/ML**    + **May or may not require additional signalling overhead for obtaining input/output data**   + **Easier LCM for gNB**   + **This can be implemented in specification transparent manner** **for network-side AI/ML model**   + **This may impose implementation restriction for using classification model** * **Alt.4: The L1-RSRP difference evaluated by comparing measured RSRP and predicted RSRP**   + **Small L1-RSRP difference does not mean high beam prediction accuracy**   **This imposes implementation restriction for using regression model** |
| HW/HiSi | Support |
| QC | How does this help with the progress of the SI? At least one company has provided some evaluation results which show the feasibility of such scheme. Why not look at it more closely during Rel-19 WI and see if it is indeed helpful? |
| Lenovo | Fine |
| NEC | Share similar view as Samsung and QC. |
| Futurewei | Support |
| Ericsson | Same view as Samsung. Good to list pros/cons for each alternative for the TR. Moreover, the specification impact is not entirely clear for alternative 3, so we propose the following amendment that can serve as a conclusion for future discussions,  ***Proposal 4.1.1: Regarding the performance metric(s) of AI/ML model monitoring for BM-Case1 and BM-Case2, conclude that potential specification impact for “Alt.3:*** ***Performance metric based on input/output data distribution of AI/ML” is more unclear in comparison to Alt1 and Alt4.*** |
| Nokia/NSB | Ok with a conclusion as suggested by E///. |
| Spreadtrum | Support the proposal. |
| MediaTek | We share similar views with Samsung and QC. |
| CMCC | Ok. |
| Fujitsu | Support the proposal |
| Google | Support |
| Mod | In order to ensure the same principle for both Proposal 4.1.1 and Proposal 3.2.1 (make down-selection or not make down-selection), A table is added to collect pros/cons.  This is the last meeting. If companies want any progress, please don’t focus on too much details and “accurate” assessment. |
| NTT DOCOMO | Since the performance monitoring is for monitoring the performance of functionality/model not system/link performance, performance indication should focus on only reflection to prediction accuracy of functionality/model.  Also, reporting of calculated KPI should be captured in Alt.1 as well as Alt.4.  Then, we propose the following update.   |  |  |  |  |  | | --- | --- | --- | --- | --- | |  | Alt.1: Beam prediction accuracy related KPIs, e.g., Top-K/1 beam prediction accuracy | Alt.2: Link quality related KPIs, e.g., throughput, L1-RSRP, L1-SINR, hypothetical BLER | Alt.3: Performance metric based on input/output data distribution of AI/ML | Alt.4: The L1-RSRP difference evaluated by comparing measured RSRP and predicted RSRP | | Performance indication | Reflect the prediction accuracy of AI model  ~~Not reflect the system/link performance directly~~ | ~~Reflect the system/link performance~~  Cannot not differentiate the impact of AI model prediction and the link quality | Not reflect the prediction performance of AI model directly  ~~Not reflect the system/link performance directly~~ | Reflect accuracy of the predicted1-RSRP  ~~Not reflect the system/link performance directly~~ | | Specification impact | Configuration/measurement/reporting for KPI calculation  Reporting of calculated KPI | Measurement/reporting (e.g., L1-RSRP) |  | Measurement/reporting of Top-K predicted beam  Reporting of predicted RSRP | |
| CMCC | It seems the table intend for NW side ***model monitoring of UE side model. With regard to*** NW side ***model monitoring of NW side model, measurement result of Set B and Top 1 beam ID need to be reported withAlt1, measurement result of Set B and Top 1 beam RSRP need to be reported withAlt4. The overhead of Alt1 and Alt4 is similar.***  ***Model monitoring type and sidedness of model need clarification in main part.*** |
| CATT | We share same view with NTT DOCOMO and propose to remove the “Not reflect the system/link performance directly”. The intention of model performance monitoring is to reflect the prediction performance of AI model rather than system/link performance. It’s very nature that the model perform monitoring metric cannot reflect system/link performance. The current version seems like the model perform metric should reflect prediction accuracy of AI model and system/link performance. Thus, it’s better to remove it from Alt.1/3/4. |
| HW/HiSi | Not support.  The details shown in this table are going ahead of the discussions that would be needed to conclude this table. For example: for spec impact reporting of KPI, this is not needed if the same entity is used for monitoring and inference. |
| Mod | * Highlight the parts that some companies think not needed and the new added part   @CMCC: For Alt.1, UE may measure set A whereas UE may measure Top-K beams for Alt.4. Thus, the RS overhead may be less for Alt.4  @Huawei: Detailed comments (e.g., how to modify the contents) are appreciated so that we can refine the table. |
| Apple | “Number of require samples”, all the proposed solutions can generate performance metric with a given number of measurement resources. Is there any analysis on the confidence interval associated with all the solutions? If not, I am afraid we are not in a position to conclude on which alternative requires more overhead. Due to that, I suggest we remove the “Number of require samples” row. |
| ZTE | We suggest to delete the rows of ‘Number of require samples’ and ‘Specification impact’ as they need more study for obtaining a conclusion, especially when we differentiate the monitoring of NW-side model and UE-side model. For the row of ‘Signaling/RS overhead’ with an UE-side model, it seems the RS overhead of Alt.2 and Alt.3 are the same. |
| Mod | The table is updated to highlight the controversial part and new added information |
| InterDigital | We are fine in principle, but prefer to delete ‘Cannot not differentiate the impact of AI model prediction and the link quality’ for Link quality related KPIs (Alt.2). This is not true. Alt.2 can differentiate the impact by having similar channels conditions for non-AI and AI. |
| CMCC | Regarding RS overhead, it seems the RS overhead of Alt.1 and Alt.4 are the same.  Average L1-RSRP difference of Top-1 predicted beam is defined as the difference between the ideal L1-RSRP of Top-1 predicted beam and the ideal L1-RSRP of the Top-1 genie-aided beam. Thus, all beams in set A need sweeping, this is same as Alt1. |
| CATT | Regarding the spec impact part, we suggest to use the similar method of Alt.1 to describe that of Alt.4. The spec impact includes configuration, measurement and reporting. And one way is the UE report the data for KPI calculation, and the network calculates the performance metric. Another method is UE calculates the performance metric and report the calculated KPI to network.  Thus ,we propose to update the spec impact of Alt.4 as following and update the table with the change mark:  Configuration/Measurement/reporting of Top-K predicted beam  Reporting of calculated KPI (if applicable) |
| New H3C | We suggest removing “Applicable to different AI models” line. Many cases require combination of several alternatives to monitor the AI performance |
| Mod | Two remaining rows for further discussion |

## NW-side model

In previous RAN1 meeting(s), the agreement(s)/conclusion(s) were made as below:

|  |
| --- |
| **RAN1#110bis-e**  Agreement  For BM-Case1 and BM-Case2 with a network-side AI/ML model, study the NW-side model monitoring:   * NW monitors the performance metric(s) and makes decision(s) of model selection/activation/ deactivation/switching/ fallback operation   Agreement  Regarding NW-side model monitoring for a network-side AI/ML model of BM-Case1 and BM-Case2, study the potential specification impacts from the following aspects   * Beam measurement and report for model monitoring * Note: This may or may not have specification impact.   **RAN1#111**  Agreement  Regarding NW-side model monitoring for a network-side AI/ML model of BM-Case1 and BM-Case2, study the necessity and the potential specification impacts from the following aspects:   * UE reporting of beam measurement(s) based on a set of beams indicated by gNB. * Signaling, e.g., RRC-based, L1-based. * Note: Performance and UE complexity, power consumption should be considered. |

The related proposals in tdocs are copied as below:

|  |  |
| --- | --- |
| Huawei[4] | *Observation 9: For AI/ML model monitoring for BM-Case1 and BM-Case2,*  *• For network side monitoring, the performance metrics for AI/ML and benchmark(s) can be based on the UE report of Opt.1(M1 L1-RSRPs and indication)/Opt.2(M2 L1-RSRPs)/Opt.3(M3 beam indices).*  *• For UE side monitoring, the performance metrics for AI/ML and genie-aided best beam can be based on the measurement of Set B and Set A, respectively.* |
| Vivo[7] | *Proposal 44: For BM-Case1 and BM-Case2 with a NW-side AI/ML model, regarding NW-side performance monitoring, study the following monitoring procedures:*  * UE performs resource measurement and reports corresponding measurement results including set B results and set A label data*  * NW performs beam prediction and predicted results comparison with label data to obtain performance metric(s)*  * NW makes decision(s) of model selection/activation/deactivation/switching/ fallback operation*  * Note: it can be applied on both model ID based and functionality-based LCM procedures*  *Proposal 45: Support to study hybrid-side model monitoring for BM-Case1 and BM-Case2 with a NW-side AI/ML model, which can save large measurement resources and report overhead compared to NW-side model monitoring.*  *Proposal 46: For BM-Case1 and BM-Case2 with a NW-side AI/ML model, regarding hybrid-side performance monitoring, study the following monitoring procedures:*  * UE performs resource measurement and reports set B results used for NW-side beam prediction*  * NW performs beam prediction based on set B results and indicates inference result (e.g., top-N predicted results) to UE*  * UE performs predicted result comparison with label data to obtain performance metric(s) and reports monitoring result(s) to gNB*  * NW makes decision(s) of model selection/activation/deactivation/switching/fallback operation*  * Note: it can be applied on both model ID based and functionality-based LCM procedures*  *Proposal 47: For BM-Case1 and BM-Case2 with a NW-side AI/ML model, study the potential specification impact on resource configuration for model monitoring:*  *• Specific beam pair resource configuration for Set B/Set C and/or Set A*  *• P3+P2 resource configuration that Rx beam assumption of P2 resource measurement is the best Rx beam searched from P3 procedure for performance improvement*  *Proposal 48: For BM-Case1 and BM-Case2 with a NW-side AI/ML model, study the potential specification impact on assistance information for model monitoring:*  *• Proprietary processed Rx beam information as assistance information from UE to NW, including measured Rx beam information, expected Rx beam information, and best Rx beam information.*  *Proposal 49: For BM-Case1 and BM-Case2 with a NW-side AI/ML model, study the potential specification impact on report overhead reduction for model monitoring:*  *• Reducing unnecessary L1-RSRP report where the omitted L1-RSRPs may be lower than a pre-defined threshold*  *Proposal 50: For BM-Case1 and BM-Case2 with a NW-side AI/ML model, study potential specification impact on quantization enhancement for model monitoring:*  *• High-precision L1-RSRP quantization*  *• Multi-resolution L1-RSRP quantization, e.g. high-resolution quantization for a group of best RSRPs and low-resolution quantization for others.* |
| Intel[9] | *Proposal 4: For a network side AI/ML model monitoring, measurement and reporting defined for data collection can be reused.* |
| Ericsson[11] | *Observation 6 System/link level performance metrics-based model monitoring method has low complexity and low signalling overhead. It can be sufficient for the NW to monitoring the AI-feature performance of users with MBB services if fall back operations are supported.*  *Observation 7 Monitoring of NW-sided models can be done in a step-wise approach by considering different performance metric(s) and associated performance monitoring related KPIs (e.g., latency, complexity, signaling overhead, etc.).* |
| CATT[15] | *Observation 1: For BM-Case1 and BM-Case2 with NW-side performance monitoring, different reporting contents from UE are needed if different kinds of beam prediction accuracy (e.g. Top-K/1 beam prediction accuracy, Top-1/K beam prediction accuracy, Top-1 beam prediction accuracy within 1 dB margin) are used as the performance metric.* |
| Nokia[19] | *Proposal 9. For NW-sided BM-Case1, the following potential specification impact can be considered,*  *• …*  *• For performance monitoring at the NW, enhancements to the CSI reporting may not be needed to enable full/partial Set A beam measurements.*  *Proposal 10. For NW-sided BM-Case2, the following potential specification impact can be considered,*  *• …*  *• For performance monitoring at the NW, enhancements to the CSI reporting may not be needed to enable full/partial Set A beam measurements.*  *Proposal 25. For NW-sided BM-case1/2, discuss signaling of configuring UE for data recording and reporting for beam measurements of Set B/A corresponding to the failure instances of the NW-sided model.* |
| Xiaomi[21] | *Proposal 13: For NW-side model monitoring for network-side AI/ML model, support to report both set B and set C, where set B will be used as network-side AI/ML model input, and set C consists of Top-K beams by UE’s measurement of set A.*  *Proposal 14: For NW-side model monitoring for network-side AI/ML model, support an event-triggered report based on comparing the indicated TCI state and the best beams obtained by measurements.* |
| Samsung[24] | *Proposal 9. For BM-Case1 with a network-side AI/ML model, for model monitoring, the following aspects should be further study:*  * UE to report the measurement results of more than 4 beams in one reporting instance*  * Assistance information (e.g., UE speed, indoor/outdoor) associated with the beam measurements*  *Proposal 10. For BM-Case1 with a network-side AI/ML model, for the alternatives for performance metric(s) of model monitoring (agreed in RAN1#112), the following observation is made:*  * Alt.1: Beam prediction accuracy related KPIs, e.g., Top-K/1 beam prediction accuracy*  * This can be done by comparing actual beam measurement and beam prediction from Set A*  * This may require additional support of L1 beam report of the measurement results of more than 4 beams in one reporting instance*  * This may require additional support of L1 beam report of the measurement results with beam index only (e.g., without the corresponding L1-RSRP)*  * This may not reflect system performance since it is an intermediate KPI*  * Alt.2: Link quality related KPIs, e.g., throughput, L1-RSRP, L1-SINR, hypothetical BLER*  * This is an indirect metric for model monitoring, which makes it prone to the potential impact of other factors (e.g., bad linkage quality)*  * This reflects system performance; however, it is hard to differentiate the poor performance is due to wrong selection of the beams(pairs) or due to other factors (e.g., bad linkage quality)*  * Alt.3: Performance metric based on input/output data distribution of AI/ML*  * May or may not require additional signalling overhead for obtaining input/output data*  * Easier LCM for gNB*  * This can be implemented in specification transparent manner for network-side AI/ML model*  * This may impose implementation restriction for using classification model*  * Alt.4: The L1-RSRP difference evaluated by comparing measured RSRP and predicted RSRP*  * Small L1-RSRP difference does not mean high beam prediction accuracy*  * This imposes implementation restriction for using regression model* |
| Lenovo[27] | *Proposal 8: For NW-side AI/ML model performance monitoring, support Tx beam repetition for the UE to report the best L1-RSRP of a Tx beam among all its Rx beams.*  *Proposal 17: Rel-17 CSI reporting framework can be reused for NW-side beam prediction by increasing the number of beams in a beam report.*  *Proposal 18: To Support NW-side AI/ML inference, the gNB can configure one or more CSI reports for the UE to report the L1-RSRPs of all the beams configured in the CMR associated with the CSI report.* |
| CAICT[28] | *Proposal 4: For BM-Case1 and BM-Case2 with a NW-side AI/ML model, overhead reduction from UE side reporting could be considered for model monitoring.* |
|  |  |

###### Mod’s assessment

**Mod’s assessment**: Some proposals are quite general and the detailed proposals are quite divergent. Moreover, most of the detailed proposals are suggested only by one company.

**Mod’s Suggestion**: No discussion in this meeting.

|  |  |
| --- | --- |
| Company | Comments |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

## UE-side model

In previous RAN1 meeting(s), the agreement(s)/conclusion(s) were made as below:

|  |
| --- |
| **RAN1#110bis-e**  Agreement  For BM-Case1 and BM-Case2 with a UE-side AI/ML model, study the following alternatives for model monitoring with potential down-selection:   * Atl1. UE-side Model monitoring   + UE monitors the performance metric(s)   + UE makes decision(s) of model selection/activation/ deactivation/switching/fallback operation * Atl2. NW-side Model monitoring   + NW monitors the performance metric(s)   + NW makes decision(s) of model selection/activation/ deactivation/switching/ fallback operation * Alt3. Hybrid model monitoring   + UE monitors the performance metric(s)   + NW makes decision(s) of model selection/activation/ deactivation/switching/ fallback operation   **RAN1#112**  Agreement  For BM-Case1 and BM-Case2 with a UE-side AI/ML model, regarding NW-side performance monitoring, study the following aspects as a starting point including the study of necessity:   * Configuration/Signaling from gNB to UE for measurement and/or reporting * UE reporting to NW (e.g., for the calculation of performance metric) * Indication from NW for UE to do LCM operations * Other aspect(s) is not precluded * Note1: At least the performance and reporting overhead of model monitoring mechanism should be considered   Agreement  For BM-Case1 and BM-Case2 with a UE-side AI/ML model, regarding UE-side performance monitoring, study the following aspects as a starting point including the study of necessity and feasibility:   * Indication/request/report from UE to gNB for performance monitoring   + Note: The indication/request/report may be not needed in some case(s) * Configuration/Signaling from gNB to UE for performance monitoring * Other aspect(s) is not precluded   **RAN1#113**  Agreement  For BM-Case1 and BM-Case2 with a UE-side AI/ML model, regarding performance monitoring, study potential spec impact(s) from the following aspects in addition to those included in previous agreements:   * Configuration/Signalling from gNB to UE for measurement and/or reporting * UE calculates performance metric(s), either reports it to NW or reports an event to NW based on the performance metric(s)   + FFS: definition of an event and the performance metric(s) used to identify it * Indication from NW for UE to do LCM operations   Agreement  For BM-Case1 and BM-Case2 with a UE-side AI/ML model, regarding performance monitoring, study the necessity and potential spec impact(s) of the mechanism that facilitate UE to detect whether the functionality/model is suitable or no longer suitable. |

The related proposals in tdocs are copied as below:

|  |  |
| --- | --- |
| H3C [1] | *Proposal 8: For BM-Case1 and BM-Case2 with a UE-side AI/ML model, regarding performance monitoring, study the following aspects as a starting point, including the study of necessity:*  *• Metric calculation method (e.g. total measured data, filtering)*  *• Metric decision method (e.g. prediction accuracy, L1-RSRP gap)* |
| Huawei[4] | *Observation 9: For AI/ML model monitoring for BM-Case1 and BM-Case2,*  *• For network side monitoring, the performance metrics for AI/ML and benchmark(s) can be based on the UE report of Opt.1(M1 L1-RSRPs and indication)/Opt.2(M2 L1-RSRPs)/Opt.3(M3 beam indices).*  *• For UE side monitoring, the performance metrics for AI/ML and genie-aided best beam can be based on the measurement of Set B and Set A, respectively.* |
| IDC[6] | *Observation 11: Definition of ‘monitoring’ in the agreement for model monitoring is not clear enough.*  *Proposal 16: Clarify the details of ‘monitoring for each alternative including UE reporting of the performance metric(s) for Alt3.*  *Proposal 19: For configuration/signaling from gNB to UE, consider configuration of monitoring RS/channel, evaluation methodology for monitoring and confirmation on UE request/trigger.*  *Proposal 20: For indication/request/report from UE to gNB, consider reporting UE monitoring result and trigger of a model recovery procedure.*  *Proposal 21: Support “Alt.3: The beam corresponding to some or all the indicated/activated TCI state(s)” with the indicated TCI state(s) not the activated TCI state(s).*  *Proposal 22: Consider Set B change request (based on performance KPIs) and AIML mode disable indication as indicators of model performance.* |
| Vivo[7] | *Proposal 51: For BM-Case1 and BM-Case2 with a UE-side AI/ML model, regarding Hybrid-side performance monitoring, study the following monitoring procedures:*  *• UE performs label data measurement, set B measurement and beam prediction and predicted result comparison with label data to obtain performance metric(s)*  *• NW makes decision(s) of model selection/activation/deactivation/switching/fallback operation*  *• Note: it can be applied on both model ID based and functionality-based LCM procedures*  *Proposal 52: For BM-Case1 and BM-Case2 with a UE-side AI/ML model, study the potential specification impact on resource configuration for model monitoring:*  *• Specific beam pair resource configuration for Set B/Set C and/or Set A*  *• P3+P2 resource configuration that Rx beam assumption of P2 resource measurement is the best Rx beam searched from P3 procedure for performance improvement*  *Proposal 53: For BM-Case1 and BM-Case2 with a UE-side AI/ML model, study the potential specification impact on assistance information for model monitoring:*  *• Proprietary processed Tx beam information as assistance information from NW to UE*  *Proposal 54: For BM-Case1 and BM-Case2 with a UE-side AI/ML model, study the potential specification impact on request signaling for model monitoring:*  *• Resource request signaling for data collection from UE to NW*  *- Beam pair resources request for model monitoring purpose including the number of requested labels, and potentially some associated triggering events to be defined*  *- P3+P2 beam sweeping resources request for model monitoring purpose including the number of requested labels, and potentially some associated triggering events to be defined*  *• Minimum resource number request for data collection from UE to NW*  *- Minimum number of requested beams for model monitoring w or w/o resource request signaling*  *- Minimum number of requested repetitions for model monitoring w or w/o resource request signaling*  *Proposal 55: For BM-Case1 and BM-Case2 with a UE-side AI/ML model, study the potential specification impact on monitoring report for model monitoring:*  *• Monitoring result report from UE to NW, including label data report or performance metric report* |
| Intel[9] | *Proposal 5: For UE-side AI/ML model, support UE-side and Hybrid model monitoring. NW-side monitoring can be further studied and used in specific cases if the model is transferred from the NW to the UE.*  *Proposal 6: For UE side AI/ML model, support UE event driven approach for model monitoring* |
| Sony[10] | *Proposal 10 : For the measurement of model monitoring in BM-Case2, the gNB can configure the RS resources based on the timestamp-related information that can be obtained from the predicted results.* |
| Ericsson[11] | *Observation 8 Potential spec impact(s) of the mechanism that facilitate UE to detect whether the functionality/model no longer suitable could comprise a request for (re)configuration based on its performance monitoring.*  *Proposal 12 For BM-Case1 and BM-Case2 with a UE-side AI/ML model, regarding performance monitoring, further study performance metrics comprising*  *a. per-sample (e.g. L1-RSRP error or correct/incorrect beam prediction),*  *b. statistical error over a certain monitoring window. For example, the 5th,10th, …, 90th percentile of the L1-RSRP prediction error*  *Observation 9 Using inference-accuracy based model monitoring for the UE-side to monitor UE-sided models can result in large RS signalling overhead and high latency.*  *Proposal 13 For BM-Case1 and BM-Case2, regarding performance monitoring, additionally study the number of samples needed to detect an outdated model for alternative 1 and 4.* |
| LG[13] | *Proposal #11: Regarding categories of performance monitoring of UE-sided model, select one of the following two approaches*  *- Approach1: clarify the definitions of the three categories from functionality-based LCM perspective.*  *- Approach2: do not define any categories, i.e. remove categories from previous agreements/TR.*  *Proposal #12: For UE-sided AI/ML model, it should be UE to calculate performance metric and monitor its performance. NW does final decision for functionality management (e.g. (de)activation, selection) but UE does decision for model management.*  *Observation #2: Dedicated signaling or procedure for UE-side performance monitoring may not be needed by considering that Set A beams could be provided based on UE capability report.* |
| CATT[14] | *Proposal 13: For BM-Case1 and BM-Case2 with a UE-side AI/ML model, for Alt.1 UE-side model monitoring, study the potential specification impacts on the following aspects:*  *• Reporting the decision of model activation/ deactivation/switching/fallback to the network*  *• Acknowledgement mechanism of model activation/ deactivation/switching/fallback from the network.* |
| Fujitsu[16] | *Proposal 8: For BM-Case1 and BM-Case2 of NW-side performance monitoring with a UE-side AI/ML model, study the additional signalling to indicate the measured and/or predicted results for UE reporting.*  *Proposal 9: For BM-Case1 and BM-Case2 with a UE-side AI/ML model, study the potential specification impacts of NW-side performance monitoring on*  * The configuration of filtering on the performance metric calculation.*  *Proposal 10: For BM-Case1 and BM-Case2 of the performance monitoring with a UE-side AI/ML model, to reduce the latency on the following LCM operations, study the feasibility on monitoring the inactive models/functionalities (e.g., the capability of UE).*  *Proposal 11: For BM-Case1 and BM-Case2 of the performance monitoring with a UE-side AI/ML model, study the potential specification impacts on* *inactive models/functionalities.*  * The configuration of performance monitoring on inactive models/functionalities*  * The reporting of monitoring results on inactive models/functionalities* |
| CMCC[17] | *Proposal 14: For BM-Case1 with a UE-side AI/ML model, study the decision report and acknowledgement mechanism for UE-side model monitoring.*  *Proposal 15: For BM-Case1 with a UE-side AI/ML model, study the information needed for UE reporting to NW to calculate the performance metric for NW-side model monitoring.* |
| Nokia[19] | *Proposal 5. For UE-sided BM-Case1 and BM-Case2, for any functionality activated towards the UE,*  *• The gNB shall be able to do the performance monitoring at the NW side.*  *• Support Alt.1: The best beam(s) obtained by measuring beams of a set indicated by gNB (e.g., Beams from Set A)*  *• To support Alt. 1, a dedicated beam measurement and reporting configuration that enables measurement and reporting of full/partial Set A (associated with a given functionality) can be used.*  *Proposal 6. For UE-sided BM-Case1 and BM-Case2, for any functionality activated towards the UE,*  *• The gNB shall be able to configure a performance monitoring KPI (e.g., Top-K/1 beam accuracy), performance monitoring resources, threshold for monitoring KPI, and monitoring window to determine functionality performance/failures of the activated functionality.*  *o Monitoring resources: Support Alt.1: The best beam(s) obtained by measuring beams of a set indicated by gNB (e.g., Beams from Set A).*  *• To enable reporting of the monitoring KPI, a dedicated beam measurement and reporting configuration that enables reporting of monitoring KPI can be used.*  *o The UE shall consider the monitoring KPI (Top-K/1 beam accuracy) with gNB configured threshold to determine functionality failures of the activated functionality.*  *• Further consider additional aspects of functionality failure detection for an activated functionality.*  *• For BM-Case2, further consider the reporting enhancements to report the KPIs based on overlapping portions of prediction and measurement windows and support methods of controlling the overlaps.* |
| Xiaomi[21] | *Proposal 10: For UE-side AI/ML model with UE-side model monitoring, support UE to indicate the decision to NW.*  *Proposal 11: For UE-side AI/ML model with NW-side and hybrid model monitoring, support an event-triggered report of performance metric from UE based on a threshold configured by gNB.*  *Proposal 12: For UE-side AI/ML model, UE-side initiated performance monitoring based on RA or SR can be considered, and NW-side initiated performance monitoring based on measurement configuration via RRC can be considered.* |
| DCM[22] | *Proposal 12: Study the L1/L2 reporting of the calculated performance metrics, event occurrence, and upcoming operation request for near real time performance monitoring.*  *Proposal 13: Prioritize hybrid performance monitoring over NW side performance monitoring due to the large overhead.* |
| OPPO[23] | *Proposal 16: For BM-Case1 and BM-Case2 with UE-side model, study the (Alt1) UE-side model monitoring as a starting point.*  *Proposal 17: For BM-Case1 and BM-Case2 with UE-side model, study whether to define to LCM-related event(s) which could be notified to upper layer of UE for LCM purpose.*  *Observation 12: For BM-Case1 and BM-Case2 with a network-side AI/ML model and monitoring, there may be no additional specification impact on LCM.* |
| Samsung[24] | *Proposal 16. For BM-Case1 with a UE-side AI/ML model, Alt2 (i.e., NW-side model monitoring) and Alt3 (i.e., Hybrid model monitoring) are preferred.*  *Proposal 18. For BM-Case1 with a UE-side AI/ML model, regarding NW-side performance monitoring, for the alternatives for performance metric(s) of model monitoring (agreed in RAN1#112), the following observation is made:*  * Alt.1: Beam prediction accuracy related KPIs, e.g., Top-K/1 beam prediction accuracy*  * This can be done by comparing actual beam measurement and beam prediction from Set A*  * This may require additional support of L1 beam report of the measurement results of more than 4 beams in one reporting instance*  * This may require additional support of L1 beam report of the measurement results with beam index only (e.g., without the corresponding L1-RSRP)*  * This may not reflect system performance at UE since it is an intermediate KPI*  * Alt.2: Link quality related KPIs, e.g., throughput, L1-RSRP, L1-SINR, hypothetical BLER*  * This is an indirect metric for model monitoring, which makes it prone to the potential impact of other factors (e.g., bad linkage quality)*  * This reflects system performance at UE; however, it is hard to differentiate the poor performance is due to wrong selection of the beams(pairs) or due to other factors (e.g., bad linkage quality)*  * Alt.3: Performance metric based on input/output data distribution of AI/ML*  * This requires gNB knowledge of input/output format of AI/ML model at UE-side*  * This can be implemented in specification transparent manner for UE-side AI/ML model*  * This may impose implementation restriction for using classification model*  * Alt.4: The L1-RSRP difference evaluated by comparing measured RSRP and predicted RSRP*  * Small L1-RSRP difference does not mean high beam prediction accuracy*  * This imposes implementation restriction for using regression model*  *Proposal 19. For BM-Case1 with a UE-side AI/ML model, regarding Hybrid performance monitoring, for the alternatives for performance metric(s) of model monitoring (agreed in RAN1#112), the following observation is made:*  * Alt.1: Beam prediction accuracy related KPIs, e.g., Top-K/1 beam prediction accuracy*  * This can be done by comparing actual beam measurement and beam prediction from Set A*  * This may not reflect system performance at UE since it is an intermediate KPI*  * Alt.2: Link quality related KPIs, e.g., throughput, L1-RSRP, L1-SINR, hypothetical BLER*  * This is an indirect metric for model monitoring, which makes it prone to the potential impact of other factors (e.g., bad linkage quality)*  * This reflects system performance at UE; however, it is hard to differentiate the poor performance is due to wrong selection of the beams(pairs) or due to other factors (e.g., bad linkage quality)*  * Alt.3: Performance metric based on input/output data distribution of AI/ML*  * Easier LCM related report for UE*  * This can be implemented in specification transparent manner for UE-side AI/ML model*  * This may impose implementation restriction for using classification model*  * Alt.4: The L1-RSRP difference evaluated by comparing measured RSRP and predicted RSRP*  * This imposes implementation restriction for using regression model* |
| ETRI[26] | *Proposal 1: For BM-Case1 and BM-Case2 with a UE-side AI/ML model, regarding hybrid performance monitoring, four events can be considered as follows:*  *- The case of the statistical representative value difference between the training dataset and test dataset exceeding a predefined threshold for a certain period of time.*  *- The case of spatially discontinuous beams being selected as the optimal beams more frequently than a pre-defined threshold for a certain period of time.*  *- The case of the difference between the maximum and K-th largest values for the probability of each beam being the best over a certain time period falling below a predefined threshold.*  *- The case of the continuous difference between predicted RSRP and measured RSRP values exceeding a predefined threshold over a certain time period.* |
| Lenovo[27] | *Proposal 9: For UE-side AI/ML inference, support aperiodic beam measurement for performance monitoring and dynamic beam updating within the beam set associated with the aperiodic trigger state for beam measurement.*  *Proposal 11: Support BFR-like AI/ML model failure report for hybrid performance monitoring for UE-side AI/ML model*  *Proposal 15: Study the mechanism for beam report associated with AI/ML inference when there is no available AI/ML model for AI/inference.*  *Proposal 16: For a beam report associated with AI/ML inference, the UE indicates that the reported beams are predicted beams or measured beams in the beam report.* |
| CAICT[28] | *Proposal 5: For UE-side AI/ML model monitoring, UE side directly monitoring (Alt.1) and hybrid monitoring (Alt.3) should be considered as baseline.* |
| KT[29] | *Proposal 4. For the model training/inference/monitoring of the UE-side AI/ML model:*  *• Study how to identify the beams within Set A which is not measured.*  *• Study how to indicate the association/mapping of beams within Set A and beams within Set B.* |
| QC[30] | *Proposal 5*  *For BM-Case1 and BM-Case2 with a UE-side AI/ML model, regarding UE-side performance monitoring, study the following signalling aspects related to configuration/signalling from gNB to UE for performance monitoring:*  *• Dedicated RS from gNB to UE for performance monitoring* |
| MTK[31] | *Proposal 25: For BM-Case1 and BM-Case2 with a UE-side AI/ML model, when functionality-based LCM is applicable, study a performance monitoring method with UE initiating LCM operation request and NW indicating/granting the corresponding activation/deactivation/fallback/switching LCM operation of AI/ML functionality to UE.*  *Proposal 26: For BM-Case1 and BM-Case2 with a UE-side AI/ML model, when functionality-based LCM is applicable, identify for each AI/ML functionality whether it is feasible for UE to initiate LCM operation requests.*  *Proposal 27: For BM-Case1 and BM-Case2 with a UE-side AI/ML model, regarding the event that UE reports for performance monitoring, defining the event at least from the following aspects,*  *• The performance metrics monitored for the event*  *• The threshold of the performance metrics for determining the occurrence of the event*  *• The number of samples of the occurrence instances required for determining the occurrence of the event, where the occurrence instances are the monitoring samples that the monitored metrics falls below a threshold*  *• The number of monitoring samples required for determining the occurrence of the event*  *• The frequency of each monitoring samples*  *Proposal 28: For BM-Case1 and BM-Case2 with a UE-side AI/ML model, model, to facilitate UE to detect a monitoring event for performance monitoring, study the necessity and potential spec impact(s) of NW signaling to UE the following aspects,*  *• The performance metrics monitored for the event*  *• The threshold of the performance metrics for determining the occurrence of the event*  *• The number of samples of the occurrence instances required for determining the occurrence of the event, where the occurrence instances are the monitoring samples that the monitored metrics falls below a threshold*  *• The number of monitoring samples required for determining the occurrence of the event*  *• The frequency of each monitoring samples*  *Proposal 29: Study the necessity and spec impact of the content of collected data that UE reports for performance monitoring for BM-Case1 and BM-Case2 with a UE-side AI/ML model, including at least the following contents,*  *• Performance metrics*  *• Beam measurements*  *• Events for monitoring* |
|  |  |

###### Mod’s assessment

**Down-selection on the model/performance monitoring mechanisms**

Based on the tdocs, there are 10 companies showing their preferences on the model/performance monitoring mechanisms for UE-side model(s):

* At least 7 companies (e.g., Intel, LGE, OPPO, CAICT, CATT, CMCC, Xiaomi) suggest to support UE-side monitoring, 3 of which (CATT, CMCC, Xiaomi) suggest NW to control the LCM operations (e.g., model activation/deactivation).
* At least 4 companies (e.g., Intel, DCM, Samsung, CAICT) suggest support hybrid model monitoring.
* At least 2 companies (e.g., Nokia, Samsung) suggest to support NW side model/performance monitoring
* **Mod’s assessment:** To moderator’s best knowledge, the same company may have different preference on the monitoring mechanism for Model-ID-based and functionality-based LCM framework. Thus, **it is more efficient to discuss the monitoring mechanism for these two types of LCM framework separately.** Based on the tdocs and previous discussions, moderator feels that most companies support the NW to make decision on LCM operations for functionality-based LCM framework.
* **Mod’s suggestion:** A proposal is suggested for functionality-based LCM framework for further discussion.
* **Related proposals in tdocs**
  + Intel: Proposal 5
  + LG: Proposal 12
  + OPPO: Proposal 16
  + CAICT: Proposal 5
  + CATT: Proposal 13
  + CMCC: Proposal 14
  + Xiaomi: Proposal 10
  + DCM: Proposal 13
  + Samsung: Proposal 16
  + Nokia: Proposal 5, 6

**RS resource configuration for performance monitoring**

At least, 5 companies (e.g., IDC, vivo, Sony, Nokia, QC) are discussing the configurations of RS resource, but some of the proposals are generic and lack details

* **Mod’s assessment:** The detailed configurations/signaling seems more suitable for WI.
* **Mod’s suggestion:** Discuss the detailed configuration/signaling later or in R19 WI (if any)
* **Related proposals in tdocs**
  + IDC: Proposal 19
  + Vivo: Proposal 52,
  + Sony: Proposal 10
  + Nokia: Proposal 5, 6
  + QC: Proposal 5

At least, 3 companies (e.g., 3 companies) are discussing how to calculate the metric(s), e.g., filtering method, how many samples and so on:

* **Mod’s assessment:** We have several alternatives for the performance metrics and the down-selection haven’t done so far. It would be more efficient for the group discuss detailed calculation method(s) after the group down-select one or more metrics. Moreover, these details are not critical/urgent for study item.
* **Mod’s suggestion:** The detailed calculation method(s) can be discussed later or in R19 WI (if any)
* **Related proposals in tdocs**
  + H3C: Proposal 8
  + Ericsson: Proposal 12
  + Fujitsu: Proposal 9

**Other issues**

There are also some proposals for various other aspects (e.g., definition of Event, monitoring of inactive model/functionality, beam-failure-recover-like procedure, and so on)

* **Mod’s assessment:** Most of these proposals are only suggested by limited companies. Thus, from moderator’s perspective, it is difficult to formulate some stable proposal(s).
* **Mod’s suggestion:** No discussion until there are more supporting companies.

###### Proposal 4.3.1 Closed

***Proposal 4.3.1:******For BM-Case1 and BM-Case2 with a UE-side AI/ML model, regarding UE-side performance monitoring for a functionality, specification impact is identified to support NW to indicate UE to do LCM operations at functionality level.***

|  |  |
| --- | --- |
| Company | Comments |
| Mod | * The intension of this proposal is to say that NW always make the final decision on the functionality activation/inactivation no matter which performance monitoring mechanism(s) is finally specified (i.e., UE-side, NW-side, “Hybrid”) * The monitoring mechanism for these two types of LCM framework will be discussed separately. |
| LG | Intention seems ok but we are not sure what spec enhancement it refers to. Functionality control by NW such as functionality activation/deactivation/indication could be done with legacy way, e.g. configuring a feature by RRC, (de-)activating a feature by MAC-CE, etc. Thus, this proposal may not be needed, and if really needed, it is better to be handled in framework agenda. |
| Xiaomi | There was an agreement can be seen as below. Does it mean the previous one was agreed to study, but this one means enhancement is needed after study?  For UE-side monitoring for UE-side model, does this proposal mean UE will report the decision to NW? in our point of view, this is the main difference from the previous agreement.  Agreement  For BM-Case1 and BM-Case2 with a UE-side AI/ML model, regarding performance monitoring, study potential spec impact(s) from the following aspects in addition to those included in previous agreements:   * Configuration/Signalling from gNB to UE for measurement and/or reporting * UE calculates performance metric(s), either reports it to NW or reports an event to NW based on the performance metric(s)   + FFS: definition of an event and the performance metric(s) used to identify it * Indication from NW for UE to do LCM operations |
| NTT DOCOMO | Support the principle. We suggest the following modification for the clarification.  ***For BM-Case1 and BM-Case2 with a UE-side AI/ML model, regarding performance monitoring, if functionality-based LCM is applicable, enhance specification to support the indication from NW for UE to do LCM operations at functionality scale*** |
| CATT | We think for UE-side model with both functionality-based LCM and model ID-based LCM, the final decision is made by NW. Otherwise, the network performance may be impacted by UE’s decision. . |
| ZTE | Support. For the differentiation of functionality-based LCM and model-ID-based LCM, whatever we call it functionality, model or just conditions, it’s just a function unit at UE that is not transparent to network and thus an unified signaling framework shall be applied. |
| Samsung | We are fine with the case that gNB controls LCM operations for functionality-based model.  If we want to move a step forward, this proposal is better to be an outcome of study. Hence, some change suggestions are provided as follows for better reflecting this intention.  ***Proposal 4.3.1:******For BM-Case1 and BM-Case2 with a UE-side AI/ML model, regarding performance monitoring, if functionality-based LCM is applicable, ~~enhance~~ consider the following specification enhancement:***   * ***~~to support the~~ indication from NW for UE to do LCM operations*** |
| HW/HiSi | [Not support].  Some clarification is needed. For functionality based LCM, at least to some extent it should be possible in a NW- transparent way. |
| QC | Share similar view as LG. There’s no beam management-specific aspect in this proposal, and hence, if further enhancement is needed, it can be dealt with in 9.2.1. |
| Lenovo | The enhancement point is not clear. How about the following update:  ***Proposal 4.3.1:******For BM-Case1 and BM-Case2 with a UE-side AI/ML model, regarding performance monitoring, if functionality-based LCM is applicable, ~~enhance specification~~ ~~to~~ support the indication from NW for UE to do LCM operations*** |
| Ericsson | Not support.  Same comment as LG, QC. There should be some BM-specific aspects on what such indication can compromise for example. |
| Nokia/NSB | “***if functionality-based LCM is applicable***” sounds bit weird as the finality LCM shall always be applicable. |
| Mod | The proposal is updated |
| CMCC | Ok. |
| Fujitsu | We are fine with the case that gNB controls LCM operations. But it’s better to have some BM-specific aspects, otherwise, it’s the scope for 9.2.1. |
| Google | We also think this belongs to 9.2.1. |
| Mod | We have three agreement for performance monitoring. For two of them, we have the bullet   * Indication from NW for UE to do LCM operations   For the UE-side performance monitoring, we don’t have the above bullet. If the LCM is done at the functionality level, most companies think NW makes the final decision. That is why the proposal is suggested.  In AI 9.2.1, there are different alternatives. This proposal is to do some down-selection for BM cases. |
| Xiaomi | Fine |
| CMCC | Ok. |
| NEC | OK. |
| Mod | Some agreement in AI 9.2.1 can cover this one. |

# Functionality/Model-ID based LCM

In previous RAN1 meeting(s), the agreement(s)/conclusion(s) were made as below:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **RAN1#111**  Working Assumption (AI 9.2.1)   |  |  | | --- | --- | | Terminology | Description | | Model identification | A process/method of identifying an AI/ML model for the common understanding between the NW and the UE  Note: The process/method of model identification may or may not be applicable.  Note: Information regarding the AI/ML model may be shared during model identification. |  |  |  | | --- | --- | | Terminology | Description | | Functionality identification | A process/method of identifying an AI/ML functionality for the common understanding between the NW and the UE  Note: Information regarding the AI/ML functionality may be shared during functionality identification.  FFS: granularity of functionality |   Note: whether and how to indicate Functionality will be discussed separately.  Agreement(AI 9.2.1)  For UE-part/UE-side models, study the following mechanisms for LCM procedures:   * For functionality-based LCM procedure: indication of activation/deactivation/switching/fallback based on individual AI/ML functionality   + Note: UE may have one AI/ML model for the functionality, or UE may have multiple AI/ML models for the functionality.   + FFS: Whether or how to indicate functionality * For model-ID-based LCM procedure, indication of model selection/activation/deactivation/switching/fallback based on individual model IDs   **RAN1#112**  Agreement(AI 9.2.1)  For UE-side models and UE-part of two-sided models:   * For AI/ML functionality identification   + Reuse legacy 3GPP framework of Features as a starting point for discussion.   + UE indicates supported functionalities/functionality for a given sub-use-case.     - UE capability reporting is taken as starting point. * For AI/ML model identification   + Models are identified by model ID at the Network. UE indicates supported AI/ML models. * In functionality-based LCM   + Network indicates activation/deactivation/fallback/switching of AI/ML functionality via 3GPP signaling (e.g., RRC, MAC-CE, DCI).   + Models may not be identified at the Network, and UE may perform model-level LCM.     - Study whether and how much awareness/interaction NW should have about model-level LCM * In model-ID-based LCM, models are identified at the Network, and Network/UE may activate/deactivate/select/switch individual AI/ML models via model ID.   FFS: Relationship between functionality identification and model identification  FFS: Performance monitoring and RAN4 impact  FFS: detailed understanding on model  Agreement(AI 9.2.1)   * AI/ML-enabled Feature refers to a Feature where AI/ML may be used.   Agreement(AI 9.2.1)   * For functionality identification, there may be either one or more than one Functionalities defined within an AI/ML-enabled feature.   **RAN1#112bis-e**  Agreement (AI 9.2.1)   * For AI/ML functionality identification and functionality-based LCM of UE-side models and/or UE-part of two-sided models:   + Functionality refers to an AI/ML-enabled Feature/FG enabled by configuration(s), where configuration(s) is(are) supported based on conditions indicated by UE capability.   + Correspondingly, functionality-based LCM operates based on, at least, one configuration of AI/ML-enabled Feature/FG or specific configurations of an AI/ML-enabled Feature/FG.     - FFS: Signaling to support functionality-based LCM operations, e.g., to activate/deactivate/fallback/switch AI/ML functionalities     - FFS: Whether/how to address additional conditions (e.g., scenarios, sites, and datasets) to aid UE-side transparent model operations (without model identification) at the Functionality level     - FFS: Other aspects that may constitute Functionality   + FFS: which aspects should be specified as conditions of a Feature/FG available for functionality will be discussed in each sub-use-case agenda. * For AI/ML model identification and model-ID-based LCM of UE-side models and/or UE-part of two-sided models:   + model-ID-based LCM operates based on identified models, where a model may be associated with specific configurations/conditions associated with UE capability of an AI/ML-enabled Feature/FG and additional conditions (e.g., scenarios, sites, and datasets) as determined/identified between UE-side and NW-side.   + FFS: Which aspects should be considered as additional conditions, and how to include them into model description information during model identification will be discussed in each sub-use-case agenda.   + FFS: Relationship between functionality and model, e.g., whether a model may be identified referring to functionality(s).   + FFS: relationship between functionality-based LCM and model-ID-based LCM * Note: Applicability of functionality-based LCM and model-ID-based LCM is a separate discussion.   **RAN1#113**  Agreement  For BM-Case1 and BM-Case2 with a UE-side AI/ML model, study the necessity and potential BM-specific conditions/additional conditions for functionality(ies) and/or model(s) at least from the following aspects:   * information regarding model inference * Set A / Set B configuration * performance monitoring * data collection * assistance information   Agreement(AI 9.2.1)  For model identification of UE-side or UE-part of two-sided models, categorize model identification types as follows, and further study relevant aspects, necessity, and specification impact (if any).   * Type A: Model is identified to NW (if applicable) and UE (if applicable) without over-the-air signaling   + The model may be assigned with a model ID during the model identification, which may be referred/used in over-the-air signaling after model identification.   + FFS: Spec impact to other WGs * Type B: Model is identified via over-the-air signaling,   + Type B1:     - Model identification initiated by the UE, and NW assists the remaining steps (if any) of the model identification       * the model may be assigned with a model ID during the model identification     - FFS: details of steps   + Type B2:     - Model identification initiated by the NW, and UE responds (if applicable) for the remaining steps (if any) of the model identification       * the model may be assigned with a model ID during the model identification     - FFS: details of steps * Note: The support and applicability of each model identification Type is a separate discussion. This study does not imply that model identification is necessary.   Agreement (AI 9.2.1)  For functionality/model-ID based LCM,   * Once functionalities/models are identified, the same or similar procedures may be used for their activation, deactivation, switching, fallback, and monitoring.   Agreement (AI 9.2.1)  Once models are identified, UE can indicate supported AI/ML model IDs for a given AI/ML-enabled Feature/FG in a UE capability report as starting point.   * FFS: applicability to model identification, Type A, type B1 and type B2   + FFS: Using a procedure other than UE capability report * Note: model identification using capability report is not precluded for type B1 and type B2 |

The related proposals in tdocs are copied as below:

|  |  |
| --- | --- |
| Futurewei[2] | *Proposal 6: Regarding AI/ML-based beam management, study the standards impact, including AI/ML related UE configuration/capability reporting, which may be used in the LCM like AI/ML model selection/configuration (like activation/deactivation) in case multiple trained AI/ML models are deployed, or other LCM procedures.* |
| NVIDIA[3] | *Proposal 10: For AI/ML based beam prediction in spatial/time domain, study potential specification impact related to UE capability for AI/ML based beam prediction including model training, model inference and model monitoring.*  *Proposal 11: For AI/ML based beam prediction in spatial/time domain, study the aspects that should be specified as conditions of a Feature/FG available for functionality.*  *Proposal 12: For AI/ML based beam prediction in spatial/time domain, study the aspects that should be considered as additional conditions and how to include them into model description information during model identification.* |
| Huawei[4] | *Proposal 18: Functionality based LCM is appropriate for UE-side model of BM-Case 1/2.*  *• Note: Whether model based LCM is applicable to UE-side model can be further clarified.*  *Proposal 19: Regarding the conditions for functionality(ies), discuss the configuration of RS configuration and CSI reporting with high priority.*  *• Additional conditions can be discussed with lower priority after its content, necessity and feasibility are clarified.* |
| Vivo[7] | *Proposal 3: For the sub use case BM-Case1 and BM-Case2, at least support following static applicable conditions,*  *• Conditions on the AI/ML model input*  * Input size, input type*  *• Conditions on information that can be derived based on AL/ML model output*  * Output size, output type*  *• Conditions on Set B*  * Pattern type, Set B size, time-domain pattern*  *• Conditions on Set A*  * Set A size*  *• Conditions on the relationship of Set B and Set A*  * Subset, different*  *• Conditions on performance monitoring*  * Monitoring type, monitoring report type, monitoring periodicity*  *Proposal 4: For the sub use case BM-Case1 and BM-Case2, at least support following additional conditions,*  *• Conditions on data collection*  * Site/Scenarios/Dataset related information, e.g. Dataset ID*  *Proposal 17: Both model ID-based LCM and functionality-based LCM should be studied for beam management.* |
| Ericsson[11] | *Proposal 14 Consider Table 2 as a starting point for outlining conditions for the BM use case,*  *Proposal 15 Study which of the outlined conditions for UE-sided AI/ML beam prediction functionality that can be part of UE capability (static), and outside of UE capability (dynamic)*   |  |  | | --- | --- | | LCM aspect | Conditions | | Information regarding model inference (what is needed to execute the model)   * Set A/B configuration, * Assistance info * UE conditions to be fulfilled | * Supported frequency-layers * Supported set A/B dimension, e.g [4,8,16,32,64] * Supported T1/T2 assumptions for case-2 * Supported TX beam IDs * Supported cell IDs * Supported UE mobility * Supported UE locations   …. | | Data collection | * Requested beams to be measured * Requested Periodicity | | Model monitoring | * Requested beams to be measured for monitoring * Requested monitoring periodicity |   **Table 2: Conditions for the BM use case** |
| Google[12] | *Proposal 8: Support the UE reports the preferred SSB/CSI-RS configuration for beam prediction including at least the preferred intervals between every two consecutive SSB/CSI-RS instances and minimum number of SSB/CSI-RS instances for CSI prediction.* |
| CATT[14] | *Proposal 8: For BM-Case1 and BM-Case2 with a UE-side AI/ML model, study the necessity and potential BM-specific conditions/additional conditions for functionality(ies) and/or model(s) at least from the following aspects:*  * Model inputs, e.g., the Set B pattern information, the periodicity and measurement time instances of T1 for BM-Case2*  * Model outputs, e.g., the model outputs can be Top-K predicted beams (pairs) or predicted L1-RSRP, the periodicity and prediction time instances of T2 for BM-Case2*  * Model performance metric, e.g., beam prediction accuracy, or predicted L1-RSRP difference*  *Proposal 9: Regarding the functionality identification of BM-Case1 and BM-Case2, study distinguishing different functionalities by large granularity characteristics, e.g. input type and output type.* |
| Nokia[19] | *Proposal 1. For UE-sided BM-Case1, RAN1 to support further details on conditions for functionalities,*  *• Supported beam prediction mode (e.g., Top-1/2/4/8 DL Tx beam prediction)*  *• Set B conditions (e.g., Measured DL RS (SSB, CSI-RS), Measured DL RS set dimension (4, 8, 12, [16]), Measured DL RS set pattern)*  *• Set A conditions (e.g., Predicted DL RS set dimension (16, 32, 64))*  *• NW-sided performance monitoring conditions (e.g., support measurements of Predicted DL RS set (full Set A, partial Set A), Measurement periodicity (100 ms, 200 ms))*  *• Conditions on supporting ML functionalities (e.g., Max number of supported functionalities (1, 2, 4, 8,.), identical or different delay on activating a functionality (2 ms, 4 ms), Generalization condition of functionalities (yes, no))*  *Proposal 2. For UE-sided BM-Case2, RAN1 to support at least the following conditions for functionalities,*  *• Supported beam prediction mode (e.g., Top-1/2/4/8 DL Tx beam prediction)*  *• Set B conditions (e.g., Measured DL RS (SSB, CSI-RS), Measured DL RS set dimension (4, 8, 12, [16]), Measured DL RS set pattern)*  *• Set A conditions (e.g., Predicted DL RS set – number of future instances (40ms, 80ms))*  *• NW-sided performance monitoring conditions (e.g., support measurements of Predicted DL RS set (full Set A, partial Set A), Measurement periodicity (100 ms, 200 ms))*  *• Conditions on supporting ML functionalities (e.g., Max number of supported functionalities (1, 2, 4, 8,.), identical or different delay on activating a functionality (2 ms, 4 ms), Generalization condition of functionalities (yes, no))*  *Proposal 3. For UE-sided BM-Case1 and BM-Case2, RAN1 to support the following optional conditions for functionalities,*  *• Conditions for UE-sided performance monitoring*  *• Conditions for data collection (including any related assistance information)*  *• Conditions for predicted L1-RSRP and other metrics*  *• For BM-Case 2, conditions for single TCI indication (single and multi beams)*  *Proposal 4. For UE-sided BM-Case1 and BM-Case2, after functionality identification, support* *UE reporting applicable functionalities among the configured functionalities.*  *Observation 1: For UE-sided BM-Case1 and BM-Case2, identifying additional conditions can still be implicitly handled by UE reporting applicable functionalities among the configured functionalities.*  *Observation 2: For UE-sided BM-Case1 and BM-Case2, with model identification (Type A or Type B), gNB and UE can identify additional conditions associated with a UE-sided model.*  *Observation 3: For UE-sided BM-Case1 and BM-Case2, based on the latest agreement in RAN1 #113 meeting [3], it may be possible to use offline model identification (Type A) with reporting supported model-ID(s) in the UE-capability to identify additional conditions associated with a UE-sided model.*  *Proposal 8. For UE-sided BM-Case1 and BM-Case2, related to the functionality/model identification and handling of additional conditions, RAN1 shall consider the following,*  *• Functionality identification and functionality-LCM procedures are mandatory to support the beam prediction use case.*  *• Additionally, any of the following method(s) can be used to handle additional conditions associated with the functionalities,*  *o UE reporting applicable functionalities among the configured functionalities (also in Proposal 3).*  *o Offline model identification and corresponding model-ID(s) reported in UE-capability (already agreed in AI 9.2.1)*  *o Further study on how to support online model identification.* |
| Xiaomi[21] | *Proposal 17: BM Case 1 and BM Case 2 can be considered as different feature.*  *Proposal 18: Different functionality can be defined for different relationship between set B and set A.*  *• Alt.1: Set A and Set B are different (Set B is NOT a subset of Set A)*  *• Alt.2: Set B is a subset of Set A (Set A and Set B are not the same)*  *• Alt.3: Set A and Set B are the same (for BM Case 2 only)*  *Proposal 19: UE need to indicate the number of supported predicted future time instance.*  *Proposal 20: Different functionality can be defined for different repeat window for BM Case 2.*  *Proposal 21: Study assistance information from gNB to UE for model switching between different models within a same functionality. E.g.,*  * Deployment scenario*  * ISD*  * Beam codebook* |
| DCM[22] | *Proposal 10: Study what aspects should be included in the condition indicated by UE capability for UE side beam prediction. At least the following information can be considered as potential conditions.*  *・ RS configuration of Set C, e.g., numerologies, carrier frequency, bandwidth, frequency density, and number of antenna ports*  *・ Applicable NW deployment, e.g., gNB antenna/beam configuration, gNB antenna radiation pattern*  *・ Reportable information, e.g., Set A information, predicted values (predicted L1-RSRP or top-K predicted beam indices), and predicted time offset*  *・ Applicable channel property e.g., received signal strength (RSRP/SINR), interference signal strength, LOS/NLOS condition, doppler information (UE speed)* |
| MTK[31] | *Proposal 30: For BM-Case1 and BM-Case2 with a UE-side AI/ML model, for the BM-specific conditions regarding “information regarding model inference”, study the necessity at least for the following sub-conditions,*  *• conditions on the number of predicted best beams (e.g., value of K for Top-K predicted beams)*  *• conditions on the model output (e.g., predicted beam ID/confidence score of each beam/beam RSRP)*  *• conditions on the available Rx beam assumptions*  *Proposal 31: For BM-Case1 and BM-Case2 with a UE-side AI/ML model, for the BM-specific conditions regarding “performance monitoring”, study the necessity at least for the following sub-conditions,*  *• conditions on performance metrics*  *• conditions on the detectable events*  *Proposal 32: For BM-Case1 and BM-Case2 with a NW-side AI/ML model, study the necessity and potential BM-specific conditions/additional conditions of UE for functionality(ies) and/or model(s) at least from the following aspects:*  *• data collection (e.g., conditions on data omitting/filtering methods, conditions on the reporting contents)*  *• assistance information* |
|  |  |
|  |  |

###### Mod’s assessment

**Details on the conditions**

At least 9 companies (e.g., Huawei, vivo, Ericsson, Google, CATT, Nokia, Xiaomi, DCM, MTK) are discussing the details of the conditions reported by UE capability.

* **Mod’s assessment:** There are many aspects involved in the discussion of conditions, and the proposals are quite divergent. On the other side, some companies commented that this issue belongs to UE capability discussion and should be discussed in WI. Based on tdocs and previous discussion, it seems difficult to achieve consensus on the detailed conditions on top of our previous agreement.
* **Mod’s suggestion:** The BM-specific conditions can be discussed later or in R19 WI (if any).
* **Related proposals in tdocs**
  + Huawei: Proposal 19
  + Vivo: Proposal 3
  + Ericsson: Proposal 15
  + Google: Proposal 8
  + CATT: Proposal 8
  + Nokia: Proposal 1, 2, 3
  + Xiaomi: Proposal 18, 19, 20
  + DCM: Proposal 10
  + MTK: Proposal 30, 31, 32

**Details on additional conditions**

At least 3 companies (e.g., Huawei, vivo, Nokia) are discussing the issue of “additional conditions”, of which 2 companies (Huawei, Nokia) seem not supportive of additional conditions and 1 company (vivo) proposes some detailed additional conditions.

* **Mod’s assessment:** There are some parallel discussions on additional conditions in AI 9.2.1 and RAN2. Moreover, this issue is quite controversial there. Thus, moderator feels it is premature to discuss the details for BM-specific additional conditions.
* **Mod’s suggestion:** The BM-specific additional conditions can be discussed later or in R19 WI (if any).
* **Related proposals in tdocs**
  + Huawei: Proposal 19
  + Nokia: Proposal 8
  + Vivo: Proposal 4

**“Dynamic” capability reporting**

At least 2 companies (e.g., Ericsson, Nokia) are discussing what conditions can be reported “dynamically”, or UE reporting the applicable functionality among the configured functionalities.

* **Mod’s assessment:** This issue is discussed in AI 9.2.1 and RAN2, and the consensus seems not clear so far. It is better for the group to study BM-specific aspects after AI 9.2.1/RAN2 get common understanding and make consensus on the general concept/mechanism for this issue.
* **Mod’s suggestion:** Postpone the discussion on BM-specific aspects until AI 9.2.1/RAN2 make more progress on this issue.
* **Related proposals in tdocs**
  + Ericsson: Proposal 15
  + Nokia: Proposal 4

|  |  |
| --- | --- |
| Company | Comments |
|  |  |
|  |  |
|  |  |
|  |  |

# Assistance information

Assistance information may be used for AI model training, inference and/or monitoring. In previous RAN1 meeting(s), the related agreement(s)/conclusion(s) were made as below:

|  |
| --- |
| **RAN1#112**  Conclusion  Regarding the explicit assistance information from UE to network for NW-side AI/ML model, RAN1 has no consensus to support the following information   * UE location * UE moving direction * UE Rx beam shape/direction   Conclusion  Regarding the explicit assistance information from network to UE for UE-side AI/ML model, RAN1 has no consensus to support the following information   * NW-side beam shape information   + E.g., 3dB beamwidth, beam boresight directions, beam shape, Tx beam angle, etc. * Note: Other information (e.g., relative information) of Tx beam(s) preserving sensitive proprietary information is a separate discussion   + e.g., some information following the same principle of Rel-17 positioning agreement   **RAN1#113**  Agreement  Regarding data collection for BM-Case1 and BM-Case2 with a UE-side AI/ML model, study the benefits, necessity and potential specification impact of the following aspect on top of those we have agreed in previous meeting:   * Assistance information from NW to UE for UE data collection for categorizing the data for the purpose of differentiating characteristics of data   + The assistance information should preserve privacy/proprietary information. |

The related proposals in tdocs are copied as below:

|  |  |
| --- | --- |
| FUTUREWEI[2] | *Observation 1: Assistance information may come with additional cost like signaling overhead, extra UE measurement overhead (including complexity, power consumption, etc.). There is usually a trade-off between performance gain and the associated overhead. On another aspect, some of the proposed assistance information may be proprietary so neither the NW nor the UE is willing to expose it to the other side, unless there is a substantial gain for exposing such information.*  *Proposal 3: When assistance information is used as input, companies should compare performance gain and additional overhead incurred with the baseline in which the assistance information is not used.* |
| NVIDIA[3] | *Proposal 2: Comprehensive evaluation results showing convincing performance gains is needed to nail down the essential assistance information needed for the spatial-domain DL beam prediction.*  *Proposal 4: Comprehensive evaluation results showing convincing performance gains is needed to nail down the essential assistance information needed for the temporal DL beam prediction.* |
| Huawei[4] | *Proposal 5: For the study of AI/ML model input for BM-Case 1 and BM-Case 2, Alt.2, i.e. the usage of assistance information (regardless of explicit or implicit), is considered with lower priority.*  *Observation 8: The motivation of introducing the assistance information for assisting UE/network side data categorization is not clear considering the following points:*  *• UE/network can train a generalized model that is applicable to multiple scenarios/antenna layouts/beam shapes.*  *• UE/gNB can autonomously sense the scenario without the need for gNB/UE notification.*  *• The categorization principle and granularity of the scenarios identified by network/UE may not match the categorization principle of the UE/network side.*  *o To achieve aligned categorization principle, offline interoperation to the physical meaning of the scenarios/antenna layouts/beam shapes the between network side and UE side may be inevitable, which may probably disclose the proprietary.* |
| Vivo[7] | *Proposal 6: At least support Tx/Rx beam angle/ID information as assistance information for performance improvement for both BM-Case1 and BM-Case2. Other assistance information can be FFS.*  *Proposal 7: For the determination/selection of assistance information,*  *• The performance, model generalization and potential specification impacts should be considered.*  *• Study how to protect sensitive proprietary/privacy information and disclose beam specific related assistance information.*  *Proposal 8: Support proprietary protection mechanism for proprietary/privacy information disclosing issue. Detailed proprietary protection mechanism can be FFS.*  *Proposal 9: Suggest to use proprietary processed assistance information as model input to address performance deterioration and sensitive proprietary information disclosure issues in both BM-Case1 and BM-Case2, where a same mapping function is maintained for training and inference.*  *Proposal 29: For BM-Case1 and BM-Case2 with a network-side AI/ML model, study potential specification impact on assistance information for AI/ML model inference:*  *• Proprietary processed Rx beam information as assistance information from UE to NW, including measured Rx beam information, expected Rx beam information, and best Rx beam information* |
| Sony[10] | *Proposal 2 : For NW-side model, support to further study a general AI/ML model for the UEs with different capabilities.*  *Proposal 3 : For NW-side model, support to further study the UEs with different capabilities collaborative reporting assistance information to NW.* |
| LG[13] | *Proposal #1: For the UE AI/ML model training and inference, assist information on relation/association between Set A beams and Set B beams should be provided to UE. To represent beams in Set A and/or Set B while preserving sensitive proprietary information, consider following exemplary methods.*  *- Set A beams are represented by LC coefficients of Set B beams*  *- Tx beam directions are represented as ordered numbers on a 2D or 3D coordinate* |
| CATT[14] | *Proposal 5: For the AI/ML mode input for BM-Case1 and BM-Case2, suggest to further study using assistance information as model input when considering the generalization over different antenna/beam patterns.* |
| NEC[15] | *Proposal 1: Support angle related information (e.g., beam angle information, UE direction/orientation information) and positioning related information (e.g., UE position) as assistance information.*  *Proposal 2: For avoiding the proprietary/privacy of the angle related information, study to provide the assistance information (e.g., angle related information) implicitly from one side to the other side.* |
| Panasonic[18] | *Proposal 1: Feature Lead to use a separate section to discuss assistance information to address the following topics:*  *- the purpose of assistance information (Some examples listed below)*  *o Help UE to identify the scenarios/configurations*  *o Help UE to categorize the data for the purpose of differentiating characteristics of the data*  *o Improve beam prediction performance*  *- what information (Some examples listed below)*  *o gNB beam configuration/deployment ID*  *o dataset/logical model ID*  *o mapping/association of beams within Set A and beams within Set B*  * QCL relationship*  * Spatial relative relation using beam grid* |
| Nokia[19] | *Observation 6. Information about NW physical properties, e.g., NW beam codebook ID, NW antenna configuration ID identifying some information about environment category or interpreting assistance information in the format of vendor specific ID (e.g., NW codebook ID) are not considered as assistance information targeting preserves privacy/proprietary information.*  *Proposal 16. Regarding data collection for BM-Case1 and BM-Case2 for UE-side model trained at UE side, RAN1 to consider the specification impact related to NW signalling an identifier of NW configuration and/or some characteristics of the L1-RSRP measurements associated with the current NW configuration (used by the NW for CSI-RS transmission) to help UE on categorizing data for data collection.* |
| DCM[22] | *Proposal 8: Study the potential specification impacts of assistance information (e.g., virtualized ID) to help the applicability determination of NW side models. At least the following information can be included in assistance information.*  *・ Rx beam determination mechanism*  *・ UE antenna/beam configuration*  *・ UE antenna radiation pattern* |
| OPPO[23] | *Proposal 23: For the assistance information of BM-Case1 and BM-Case2, suggest to*  *• Justify the performance benefits if assistance information is used*  *• Identify whether the used assistance information would expose proprietary and/or privacy information of either NW-side or UE-side.* |
| Lenovo[27] | *Proposal 13: Study on how to obtain the assisting information for AI/ML model input for both NW-side and UE-side AI/ML inference.* |
|  |  |

###### Mod’s assessment

**Assistance information for data categorization**

In the last meeting, the group agreed to study assistance information from NW to UE for UE data collection for categorizing the data for the purpose of differentiating characteristics of data. In this meeting, at least two companies (e.g., Panasonic, Nokia) propose to use “identifier” as the assistance information to categorize the data.

Meanwhile, Huawei comments that the motivation of introducing such kind of assistance information is not clear.

* **Mod’s assessment:** Only a limited number of companies discuss this issue in the tdocs. We tried assistance information in form of “ID” in the last two meetings, but cannot achieve consensus. It seems unlikely for the group to agree to use “ID” at this meeting. On the other hand, most companies support this assistance information. Thus, it is unlikely to revert our previous agreement.
* **Mod’s suggestion:** No discussion in this meeting. Details can be discussed later or in R19 WI (if any)
* **Related proposals in tdocs**
  + Panasonic: Proposal 1
  + Nokia: Proposal 16
  + Huawei: Observation 8

**Other type of assistance information**

* Tx/Rx beam angle/ID information
  + Vivo: Proposal 6
* Tx beam directions are represented as ordered numbers on a 2D or 3D coordinate
  + LGE: Proposal 1
* Angle related information (e.g., beam angle information, UE direction/orientation information)
  + NEC: Proposal 1
* Positioning related information (e.g., UE position)
  + NEC: Proposal 1
* Rx beam determination mechanism
  + DCM: Proposal 8
* UE antenna/beam configuration
  + DCM: Proposal 8
* UE antenna radiation pattern
  + DCM: Proposal 8
* Justification is needed before the introduction of any new assistance information
  + Futurewei: Proposal 3
  + NVIDIA: Proposal 2, Proposal 4
  + Huawei: Proposal 5
  + OPPO: Proposal 23

**Mod’s assessment:** Many of the proposals are conflicting with our previous conclusions. For each type of assistance information, there is only 1 or 2 supporting companies. Moreover, there are a number of companies not support any new assistance information. Thus, it seems unlikely for the group to achieve consensus on any new type of assistance information.

**Mod’s suggestion:** No discussion in this meeting.

|  |  |
| --- | --- |
| Company | Comments |
|  |  |
|  |  |
|  |  |
|  |  |

# Other aspects of LCM / use cases

In previous RAN1 meeting(s), the agreement(s)/conclusion(s) were made as below:

|  |
| --- |
| **RAN1#110**  Agreement  Study the following aspects, including the definition of components (if needed) and necessity, in Life Cycle Management   * Data collection   + Note: This also includes associated assistance information, if applicable. * Model training * [Model registration] * Model deployment   + Note: Terminology is to be defined. ~~This includes process of compiling a trained AI/ML model and packaging it into an executable format and delivering to a target device.~~ * [Model configuration] * Model inference operation * Model selection, activation, deactivation, switching, and fallback operation   + ~~Note: some of them to be refined~~ * Model monitoring * Model update   + Note: Terminology is to be defined. This includes model finetuning, retraining, and re-development via online/offline training. * Model transfer * UE capability   Note: Some aspects in the list may not have specification impact.  Note: Aspects with square brackets are tentative ~~and pending terminology definition~~.  Note: More aspects may be added as study progresses.  **RAN1#110bis-e**  Agreement  Study LCM procedure on the basis that an AI/ML model has a model ID with associated information and/or model functionality at least for some AI/ML operations ~~when network needs to be aware of UE AI/ML models~~   * FFS: Detailed discussion of model ID with associated information and/or model functionality. * FFS: usage of model ID with associated information and/or model functionality based LCM procedure * FFS: whether support of model ID * FFS: the detailed applicable AI/ML operations   Agreement  Study various approaches for achieving good performance across different scenarios/configurations/sites, including   * Model generalization, i.e., using one model that is generalizable to different scenarios/configurations/sites * Model switching, i.e., switching among a group of models where each model is for a particular scenario/configuration/site   + [Models in a group of models may have varying model structures, share a common model structure, or partially share a common sub-structure. Models in a group of models may have different input/output format and/or different pre-/post-processing.] * Model update, i.e., using one model whose parameters are flexibly updated as the scenario/configuration/site that the device experiences changes over time. Fine-tuning is one example. |

## General views of sub use case

In previous meetings, the following agreements/conclusion were made as below:

|  |
| --- |
| **RAN1#109-e**  Agreement  For AI/ML-based beam management, support BM-Case1 and BM-Case2 for characterization and baseline performance evaluations   * BM-Case1: Spatial-domain DL beam prediction for Set A of beams based on measurement results of Set B of beams * BM-Case2: Temporal DL beam prediction for Set A of beams based on the historic measurement results of Set B of beams * FFS: details of BM-Case1 and BM-Case2 * FFS: other sub use cases   Note: For BM-Case1 and BM-Case2, Beams in Set A and Set B can be in the same Frequency Range  Conclusion  For AI/ML based beam management, RAN1 has no consensus to support on studying any other sub use case in addition to BM-Case1 and BM-Case2  Note: this conclusion is independent of the discussion on the alternatives of AI/ML model inputs for BM-Case1 and BM-Case2 |

The related proposals in tdocs are copied as below:

|  |  |
| --- | --- |
| IDC[6] | *Observation 1: The agreements made in RAN1#109 and RAN1#110bis-e do not preclude the case that beams in Set A and Set B in different frequency ranges.*  *Observation 2: Supporting Set A and Set B in different frequency ranges is beneficial considering different beamwidths especially when multiple cells in different Frequency Ranges are implemented toward to an identical direction.*  *Proposal 1: BM-Case 1 and BM-Case 2 with Set A and Set B in different frequency ranges are supported as well as in a same frequency range.*  *Proposal 8: AI/ML based beam management based on association between different frequency ranges should supported for both between FR1 and FR2-1 and between FR2-1 and FR2-2.* |
| Vivo[7] | *Proposal 1: Study the two AI-based beam prediction solutions for both BM-Case1 and BM-Case2, i.e. enhanced beam pair prediction scheme and DL Tx beam prediction scheme, and considering specification impacts with generalization aspects, such as Set B construction, supported number of Tx/Rx beams, various number of antenna configurations, etc.* |
| Sony[10] | *Proposal 5 : Support to further study the joint work of beam prediction for time domain and spatial domain.* |
| Google[12] | *Proposal 16: The study of AI/ML based BM should consider both FR1 and FR2.* |
| Samsung[24] | *Proposal 1. For BM-Case1 with a network-side AI/ML model, a typical beam prediction procedure is as follows:*  * Step#1. gNB transmits RSs corresponding to the beams from Set B L1-RSRP for all of Set B beams*  * Step#2. UE measures the RSs and provides corresponding L1-RSRP report.*  * Step#3. gNB predicts the best beam(s) in set A on the basis of the L1-RSRP report.*  * Note: Set B is a subset of Set A or Set A and Set B are different.*  *Proposal 11. For BM-Case1 with a UE-side AI/ML model, a typical beam prediction procedure is as follows:*  * Step#1. gNB transmits RSs corresponding to the beams from Set B.*  * Step#2. UE measures the RSs.*  * Step#3. UE reports the best predicted beams within Set A.*  * Note: Set B is a subset of Set A or Set A and Set B are different.*  *Proposal 20. For BM-Case2 with a network-side AI/ML model, a typical beam prediction procedure is as follows:*  * Step#1. gNB transmits RSs corresponding to the beams from Set B*  * Step#2. UE measures the RS and provides corresponding L1-RSRP report for each measurement time instance for the RS.*  * Step#3. gNB predicts the future best beam(s) in set A to the historical L1-RSRP reporting.*  * Note: Set B is a subset of Set A or Set A and Set B are different or Set A and Set B are the same.*  *Proposal 26. For BM-Case2 with a UE-side AI/ML model, a typical beam prediction procedure is as follows:*  * Step#1. gNB transmits RSs corresponding to the beams from Set B*  * Step#2. UE measures the RSs corresponding to the beams from Set B and provides the predicted/future L1-RSRP of the RSs.*  * Note: Set B is a subset of Set A or Set A and Set B are different or Set A and Set B are the same.* |
|  |  |

|  |  |
| --- | --- |
| Company | Comments |
|  |  |
|  |  |
|  |  |

## Training and deployment of AI/ML model

### Training/inference at UE/NW side (including model transfer)

In previous RAN1 meeting(s), the following agreements were made:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **RAN1#109-e**  Agreement  For the sub use case BM-Case1, consider both Alt.1 and Alt.2 for further study:   * Alt.1: AI/ML inference at NW side * Alt.2: AI/ML inference at UE side   Agreement  For the sub use case BM-Case2, consider both Alt.1 and Alt.2 for further study:   * Alt.1: AI/ML inference at NW side * Alt.2: AI/ML inference at UE side   **RAN1#110**  Agreement  At least for the sub use case BM-Case1 and BM-Case2, support both Alt.1 and Alt.2 for the study of AI/ML model training:   * Alt.1: AI/ML model training at NW side; * Alt.2: AI/ML model training at UE side.   Note: Whether it is online or offline training is a separate discussion.  Working Assumption  Include the following into a working list of terminologies to be used for RAN1 AI/ML air interface SI discussion.   |  |  | | --- | --- | | Terminology | Description | | AI/ML model delivery | A generic term referring to delivery of an AI/ML model from one entity to another entity in any manner.  Note: An entity could mean a network node/function (e.g., gNB, LMF, etc.), UE, proprietary server, etc. |   **RAN1#111**  Agreement  For the sub use case BM-Case1 and BM-Case2, at least support Alt.1 and Alt.2 for AI/ML model training and inference for further study:   * Alt.1. AI/ML model training and inference at NW side * Alt.2. AI/ML model training and inference at UE side * The discussion on Alt.3 for BM-Case1 and BM-Case2 is dependent on the conclusion/agreement of Agenda item 9.2.1 of RAN1 and/or RAN2 on whether to support model transfer for UE-side AI/ML model or not   + Alt.3. AI/ML model training at NW side, AI/ML model inference at UE side |

The related proposals in tdocs are copied as below:

|  |  |
| --- | --- |
| IDC[6] | *Observation 4: AI/ML inference/training at NW side (Alt.1) could be a good implementation option as UE implementation is generally limited due to computational power and battery consumption than gNB implementation. However, AI/ML inference/training generally requires more detailed explicit information which leads significant reporting overhead.*  *Observation 5: AI/ML inference/training at UE side (Alt.2) can be limited due to limited computational power and battery consumption at UE implementation, however, UE can easily utilize more information that the UE acquired by measuring SSB/CSI-RS without consuming any reporting overhead.*  *Proposal 5: Support both AI/ML inference/training at NW side (Alt.1) and UE side (Alt.2) for both BM-Case1 and BM-Case2.* |
| Vivo[7] | *Observation 1: Report overhead may increase dramatically but with less specification impacts for Alt. 1 with enhanced beam pair prediction solution and DL Tx beam prediction solution.*  *Observation 2: Report overhead can be reduced to top-k L1-RSRP and its related Rx beam information, but assistance information including NW-side information, such as antenna configuration, Tx beam angle, etc., should be signaled to UE for Alt.2.*  *Observation 3: Due to UE side model training, if mismatch NW-side beam information is signaled to UE, significant performance deterioration can be observed for AI based beam prediction scheme in Alt.2.*  *Observation 4: Report overhead and UE energy/complexity is limited for Alt.3, but model transfer is needed.*  *Observation 5: For Alt.3, a cell specific AI solution can be achieved with generalization consideration and infra vendor may not need to disclose NW-side information such as antenna configuration, Tx beam angle, etc.*  *Proposal 2: For the sub use case BM-Case1 and BM-Case2, support to study Alt.3 for AI/ML model training and inference:*  * Alt.3. AI/ML model training at NW side, AI/ML model inference at UE side*  *Proposal 26: For Alt.3. which is AI/ML model training at NW side and inference at UE side, it has similar data collection procedure and potential specification impacts as Alt.1, i.e. both model training and model inference at NW side, for enhanced beam pair prediction and DL Tx beam prediction scheme.*  *Proposal 39: In model inference procedure, Alt.3, i.e. model training at NW side and model inference at UE side, with enhanced beam pair prediction and DL Tx beam prediction scheme has similar specification impacts as an AI model trained and inferenced at UE side.*  *Proposal 40: Study signaling aspects enhancement related to the procedure of model transfer, model registration and model activation, for the case with AI/ML model training at NW side and AI/ML model inference at UE side.* |
| CATT[14] | *Proposal 2: For DL beam pair prediction with a UE-side model, study the following aspects:*  *• For model training at NW side, study how to report relative Rx beam information when preserving sensitive proprietary information*  *• For model training at UE side, study how to send/report relative Tx beam information when preserving sensitive proprietary information.*  *Proposal 3: For DL beam pair prediction with model training at NW side and inference at UE side, study how to align the mapping rule for Tx beam ID and Rx beam ID between the NW and UE.* |
| OPPO[23] | *Observation 1: For BM-Case1, deploying AI/ML inference at UE side can avoid beam reporting on Set B, therefore resulting in minimum standard impact.*  *Observation 2: For BM-Case2, deploying AI/ML inference at UE side seems more reasonable, otherwise (inference at NW side) there could be overwhelming beam reporting on Set B when Set B is the same as Set A.*  *Observation 3: Whether to study the case that AI/ML model for BM-Case1 and BM-Case2 trained at NW side and delivered to UE side, highly depends on the outcome of model transfer issue discussed in agenda item 9.2.1.* |
|  |  |

###### Mod’s assessment

**Mod’s assessment:** Alt.3 is depending on the discussion of model transfers in AI 9.2.1 and other WG(s) (e.g., RAN2). Based on the discussion of previous meetings, it seems unlikely to achieve any process unless other AI/WG(s) can make some progress.

**Mod’s suggestion:** Postpone the discussion to wait for more progress in other AI(s)/WG(s).

|  |  |
| --- | --- |
| Company | Comments |
|  |  |
|  |  |
|  |  |
|  |  |

### Online/offline training

In previous RAN1 meeting(s), the agreement(s)/conclusion(s) were made as below:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **RAN1#110**  Working Assumption   |  |  | | --- | --- | | Terminology | Description | | Online training | An AI/ML training process where the model being used for inference) is (typically continuously) trained in (near) real-time with the arrival of new training samples.  Note: the notion of (near) real-time vs. non real-time is context-dependent and is relative to the inference time-scale.  Note: This definition only serves as a guidance. There may be cases that may not exactly conform to this definition but could still be categorized as online training by commonly accepted conventions.  Note: Fine-tuning/re-training may be done via online or offline training. (This note could be removed when we define the term fine-tuning.) | | Offline training | An AI/ML training process where the model is trained based on collected dataset, and where the trained model is later used or delivered for inference.  Note: This definition only serves as a guidance. There may be cases that may not exactly conform to this definition but could still be categorized as offline training by commonly accepted conventions. |   Note: It is encouraged for the 3gpp discussion to proceed without waiting for online/offline training terminologies. |

The related proposals in tdocs are copied as below:

|  |  |
| --- | --- |
| NVIDIA[3] | *Observation 1: Offline training may be more feasible for the near future. But in the long run, it is vital that the AI/ML models can learn continuously to adapt to varying environments, site-specific conditions, and heterogenous configurations.* |
| IDC[6] | *Observation 22: Investigating both offline training and online training in Rel-18 is not achievable given the limited timeline for AI/ML study.*  *Observation 23: While online training requires more complicated procedures to support training in (near) real-time, offline training requires relatively simpler procedures as offline training is done by using already collected data sets.*  *Proposal 34: Prioritize offline training for the sub use case BM-Case 1 and BM-Case 2.* |
| OPPO[23] | *Proposal 1: For AI/ML beam prediction, starting from offline model training at current stage.* |
| QC[30] | *Proposal 1: For the sub use case BM-Case1 and BM-Case2 and for UE-side AI/ML models, Agenda item 9.2.3.2 should focus on offline training scenario, in which the development and training of the AI/ML model happens offline without the need to involve 3gpp signaling* |
|  |  |

###### Mod’s assessment

**Mod’s assessment:** RAN2 made agreement in RAN2#121bis to deprioritize online training. Meanwhile, based on the discussion in previous meetings, most companies suggest to prioritize offline training. Thus, there is no meaning to further discuss this issue.

|  |
| --- |
| * R2 will deprioritize aspects of on-line/real-time training for the whole SI (unless R1 identifies that it is needed for one of the studied use cases). |

**Mod’s assessment:** No discussion in this meeting.

|  |  |
| --- | --- |
| Company | Comments |
|  |  |
|  |  |
|  |  |
|  |  |

## Construction of Set A and Set B

In previous RAN1 meeting(s), the following agreements and conclusions were made:

|  |
| --- |
| **RAN1#109-e**  Conclusion  For the sub use case BM-Case1, consider the following alternatives for further study:   * Alt.1: Set B is a subset of Set A   + FFS: the number of beams in Set A and B   + FFS: how to determine Set B out of the beams in Set A (e.g., fixed pattern, random pattern, …) * Alt.2: Set A and Set B are different (e.g. Set A consists of narrow beams and Set B consists of wide beams)   + FFS: the number of beams in Set A and B   + FFS: QCL relation between beams in Set A and beams in Set B   + ~~FFS: construction of Set B (e.g., regular pre-defined codebook, codebook other than regular pre-defined one)~~ * Note1: Set A is for DL beam prediction and Set B is for DL beam measurement. * Note2: The narrow and wide beam terminology is for SI discussion only and have no specification impact * Note3: The codebook constructions of Set A and Set B can be clarified by the companies.   Conclusion  For the sub use case BM-Case2, further study the following alternatives with potential down-selection:   * Alt.1: Set A and Set B are different (e.g. Set A consists of narrow beams and Set B consists of wide beams)   + FFS: QCL relation between beams in Set A and beams in Set B * Alt.2: Set B is a subset of Set A (Set A and Set B are not the same)   + FFS: how to determine Set B out of the beams in Set A (e.g., fixed pattern, random pattern, …) * Alt.3: Set A and Set B are the same * Note1: Predicted beam(s) are selected from Set A and measured beams used as input are selected from Set B. * Note2: It is up to companies to provide other alternative(s) * Note3: The narrow and wide beam terminology is for SI discussion only and have no specification impact   **RAN1#110**  Agreement  For the sub use case BM-Case1, support the following alternatives for further study:   * Alt.1: Set A and Set B are different (Set B is NOT a subset of Set A) * Alt.2: Set B is a subset of Set A * Note1: Set A is for DL beam prediction and Set B is for DL beam measurement. * Note2: The beam patterns of Set A and Set B can be clarified by the companies.   Agreement  For the sub use case BM-Case2, further study the following alternatives:   * Alt.1: Set A and Set B are different (Set B is NOT a subset of Set A) * Alt.2: Set B is a subset of Set A (Set A and Set B are not the same) * Alt.3: Set A and Set B are the same * Note1: The beam pattern of Set A and Set B can be clarified by the companies.   Agreement (AI 9.2.3.1)   * Study the following options on the selection of Set B of beams (pairs)   + Option 1: Set B is fixed across training and inference     - FFS on the beams of Set B   + Option 2: Set B is variable (e.g., different beams (pairs) patterns in each report/measurement during training and/or inference)     - FFS on fixed or variable number of beams (pairs)     - FFS on the details   + Other options are not precluded.   + FFS on the number of beams (pairs) in Set B   **RAN1#110bis-e**  Agreement   * Study the following options on the selection of Set B of beams (pairs)   + Option 1: Set B is fixed across training and inference   + Option 2: Set B is variable (e.g., different beams (pairs) patterns in each time instance/report/measurement during training and/or inference), FFS:     - Opt A: Set B is changed following a set of pre-configured patterns     - Opt B: Set B is randomly changed among pre-configured patterns     - Opt C: Set B is randomly changed among Set A beams (pairs)     - The number of beams(pairs) in Set B can be fixed or variable     - Note: BM-Case1 and BM-Case2 may be considered for different option.   + Other options are not precluded.   Conclusion  For the sub use case BM-Case1 and BM-Case2, Set B is a set of beams whose measurements are taken as inputs of the AI/ML model, |

The related proposals in tdocs are copied as below:

|  |  |
| --- | --- |
| Huawei[4] | *Proposal 1: Study Set B is a subset of Set A and Set B is different from Set A with the same priority.*  *Proposal 4: For the study of AI/ML model input for BM-Case 1 and BM-Case 2, prioritize fixed beams.* |
| IDC[6] | *Observation 6: As using same beamwidth for all channels and signals is a general implementation within a frequency range, using a subset of Set A as Set B is a reasonable option if Set A and Set B are utilized in a same frequency range.*  *Observation 7: It is difficult to use a subset of Set A considering different beamwidths for beam management between different frequency ranges.*  *Observation 8: Utilization of wide beam information from a low frequency range has great potential as a low frequency range is more reliable and utilization of wide beam requires much less time and frequency resources for beam management.*  *Proposal 6: Support ‘Set B is a subset of Set A’ when Set A and Set B are utilized in a same frequency range for both BM-Case1 and BM-Case2.*  *Proposal 7: Support ‘Set A and Set B are different’ when Set A and Set B are utilized in different frequency ranges for both BM-Case1 and BM-Case2.* |
| Vivo[7] | *Proposal 5: Deprioritize totally random pattern in set B scheme, and support to further study specification impact on Set B with pre-configured beam patterns and Set B selected from Set C. How to select pre-configured patterns and how to configure the number of beams in Set C can be FFS.* |
| Sony[10] | *Proposal 4 : For the relationship between Set A and Set B, support both Alt.1 and Alt.2 for BM-Case1 and BM-Case2.* |
| DCM[22] | *Proposal 1: Define Set C and Set D as follows to facilitate the discussion*  *・Set C is a set of beams whose measurements are obtained by UE for model inputs (Set B ⊂ Set C)*  *・Set D is a set of beams which could be potentially measured by UE (Set C ⊂ Set D)*  *Proposal 2: Study the following scenario for the beam prediction*  *・Fixed Set C: UE measure the same beam per model inference.*  *・Variable Set C: UE may measure the different beam per model inference.* |
| OPPO[23] | *Proposal 20: For BM-Case1, Set B is a subset of Set A.*  *Proposal 21: For BM-Case2, Set B and Set A are the same.* |
|  |  |
|  |  |
|  |  |
|  |  |

###### Mod’s assessment

**Mod’s assessment**: We will focus on the spec impact (if any) of Set A/Set B in other section(s).

|  |  |
| --- | --- |
| Company | Comments |
|  |  |
|  |  |

## Input of BM-Case1 and BM-Case2

In previous RAN1 meeting(s), the agreements/conclusions were made as below:

|  |
| --- |
| **RAN1#109-e**  Conclusion  Regarding the sub use case BM-Case1, further study the following alternatives for AI/ML input:   * Alt.1: Only L1-RSRP measurement based on Set B * Alt.2: L1-RSRP measurement based on Set B and assistance information   + FFS: Assistance information. The following were mentioned by companions in the discussion:  Tx and/or Rx beam shape information (e.g., Tx and/or Rx beam pattern, Tx and/or Rx beam boresight direction (azimuth and elevation), 3dB beamwidth, etc.), expected Tx and/or Rx beam for the prediction (e.g., expected Tx and/or Rx angle, Tx and/or Rx beam ID for the prediction), UE position information, UE direction information, Tx beam usage information, UE orientation information, etc.     - Note: The provision of assistance information may be infeasible due to the concern of disclosing proprietary information to the other side. * Alt.3: CIR based on Set B * Alt.4: L1-RSRP measurement based on Set B and the corresponding DL Tx and/or Rx beam ID * Note1: It is up to companies to provide other alternative(s) including the combination of some alternatives * Note2: All the inputs are “nominal” and only for discussion purpose.   Conclusion  Regarding the sub use case BM-Case2, further study the following alternatives of measurement results for AI/ML input (for each past measurement instance):   * Alt.1: Only L1-RSRP measurement based on Set B * Alt 2: L1-RSRP measurement based on Set B and assistance information   + FFS: Assistance information. The following were mentioned by companies in the discussion:, Tx and/or Rx beam angle, position information, UE direction information, positioning-related measurement (such as Multi-RTT), expected Tx and/or Rx beam/occasion for the prediction (e.g., expected Tx and/or Rx beam angle for the prediction, expected occasions of the prediction), Tx and/or Rx beam shape information (e.g., Tx and/or Rx beam pattern, Tx and/or Rx beam boresight directions (azimuth and elevation), 3dB beamwidth, etc.) , increase ratio of L1-RSRP for best N beams, UE orientation information     - Note: The provision of assistance information may be infeasible due to the concern of disclosing proprietary information to the other side. * Alt.3: L1-RSRP measurement based on Set B and the corresponding DL Tx and/or Rx beam ID * Note1: It is up to companies to provide other alternative(s) including the combination of some alternatives * Note2: All the inputs are “nominal” and only for discussion purpose. |

The related proposals in tdocs are copied as below:

|  |  |
| --- | --- |
| H3C[1] | *Proposal4：For BM-Case1 and BM-Case2，use Alt.1(only L1-RSRP) for the study on AI/ML input with high priority.* |
| NVIDIA[3] | *Observation 2: Evaluation results show that by using L1-RSRP measurement based on Set B of beams, the AI/ML-based algorithm can achieve performance comparable to that of exhaustive beam search in Set A of beams.*  *Proposal 1: For BM-Case 1, at least support L1-RSRP measurement based on Set B of beams as AI/ML model input.*  *Observation 3: Evaluation results show that by using historical optimal index, the AI/ML-based algorithm can satisfactorily yield optimal beam index prediction for future time instances.*  *Proposal 3: For BM-Case 2 (temporal DL beam prediction), at least support using historical optimal beam index based on Set B of beams as AI/ML model input.* |
| Huawei[4] | *Proposal 2: For the remainder of the study item, do not consider CIR based on Set B as model input.*  *Proposal 3: For the study of BM-Case 1 and BM-Case 2, prioritize Alt.1 (Only L1-RSRP for Set B) as AI/ML input.*  *Observation 1: Alt.4 for the BM-Case 1 and Alt.3 for BM-Case 2 for the AI/ML model input (which are identical, both using L1-RSRP for Set B and DL Tx and/or Rx beam ID) do not provide benefits in addition to Alt.1 in case of fixed beams.* |
| IDC[6] | *Observation 9: ‘Only L1-RSRP measurement based on Set B’ is not clear enough as the alternative does not provide any beam related information.*  *• If ‘Only L1-RSRP measurement based on Set B’ means that L1-RSRP measurements are provided in a fixed order, in our view, the input is not ‘Only L1-RSRP measurement based on Set B’.*  *• Reporting L1-RSRP measurements in a fixed order is indicating L1-RSRP measurement with implicit beam related information.*  *Proposal 9: Companies supporting L1-RSRP values without beam ID should provide more details.*  *Observation 10: ‘L1-RSRP measurement based on Set B and the corresponding DL Tx and/or Rx beam ID’ can be a baseline option as AI/ML model can predict RSRP measurements with Tx and Rx beam IDs which are not provided.*  *Proposal 10: Support ‘L1-RSRP measurement based on Set B and the corresponding DL Tx and/or Rx beam ID’ as a baseline.*  *Proposal 11: Additional information such as TRP IDs and Panels IDs should be considered.*  *Proposal 12: ‘CIR based on Set B’ can be considered as an alternative only for beam management based on FR1 information.* |
| Vivo[7] | *Proposal 10: Regarding to BM-Case1 and BM-Case 2, at least prioritize following AI input information for further study on specification impact:*  *• L1-RSPR measurement based on Set B*  *• Corresponding DL Tx beam pointing angle/ID*  *• Corresponding DL Rx beam pointing angle/ID*  *• Expected Tx and/or expected Rx beam angle/ID*  *• Further discuss other information, such as Tx and/or Rx beam shape information, 3dB beam-width, etc.* |
| Google[12] | *Proposal 1: For spatial domain beam prediction, support Alt3 (CIR based on set B).*  *Proposal 2: Study the input from the beam report from a group of UEs for UE-group based beam prediction, where the UEs in a group share the similar location and velocity.* |
| CATT[14] | *Proposal 4: For the AI/ML mode input for BM-Case1 and BM-Case2, suggest to support at least following：*  *• Alt.1: Only L1-RSRP measurement based on Set B*  *• Alt.4: L1-RSRP measurement based on Set B and the corresponding DL Tx and/or Rx beam ID.* |
| Lenovo[27] | *Proposal 4: Support Alt 1 and Alt 2 as the AI/ML model output for both UE-side and NW-side inference.* |
|  |  |
|  |  |

###### Mod’s assessment

**Mod’s assessment**: We will focus on the spec impact (if any) of AI model input in other section(s).

|  |  |
| --- | --- |
| Company | Comments |
|  |  |
|  |  |

## Output of BM-Case1 and BM-Case2

In previous RAN1 meeting(s), the agreement(s)/conclusion(s) are made as below:

|  |
| --- |
| **RAN1#110**  Agreement  Regarding the sub use case BM-Case1 and BM-Case2, study the following alternatives for AI/ML output:   * Alt.1: Tx and/or Rx Beam ID(s) and/or the predicted L1-RSRP of the N predicted DL Tx and/or Rx beams   + E.g., N predicted beams can be the top-N predicted beams * Alt.2: Tx and/or Rx Beam ID(s) of the N predicted DL Tx and/or Rx beams and other information   + FFS: other information (e.g., probability for the beam to be the best beam, the associated confidence, beam application time/dwelling time, Predicted Beam failure)   + E.g., N predicted beams can be the top-N predicted beams * Alt.3: Tx and/or Rx Beam angle(s) and/or the predicted L1-RSRP of the N predicted DL Tx and/or Rx beams   + E.g., N predicted beams can be the top-N predicted beams   + FFS: details of Beam angle(s) * FFS: how to select the N DL Tx and/or Rx beams (e.g., L1-RSRP higher than a threshold, a sum probability of being the best beams higher than a threshold, RSRP corresponding to the expected Tx and/or Rx beam direction(s)) * Note1: It is up to companies to provide other alternative(s) * Note2: Beam ID is only used for discussion purpose * Note3: All the outputs are “nominal” and only for discussion purpose * Note4: Values of N is up to each company. * Note5: All of the outputs in the above alternatives may vary based on whether the AI/ML model inference is at UE side or gNB side. * Note 6: The Top-N beam IDs might have been derived via post-processing of the ML-model output |

The related proposals in tdocs are copied as below:

|  |  |
| --- | --- |
| FUTUREWEI[2] | *Proposal 4: Specify exact model outputs only when standards impact is involved while companies are encouraged to share their model output details for discussion purpose of AI/ML based beam management.* |
| Huawei[4] | *Proposal 6: For BM-Case1 and BM-Case2, in Rel-18 SI, consider Alt. 1 as the baseline for the assumption on the AI/ML model output:*   * *Alt.1: Tx and/or Rx Beam ID(s) and/or the predicted L1-RSRP of the N predicted DL Tx and/or Rx beams*    + *E.g., N predicted beams can be the top-N predicted beams* |
| IDC[6] | *Proposal 13: Support ‘Tx and/or Rx Beam ID(s) and/or the predicted L1-RSRP of the N predicted DL Tx and/or Rx beams’ as a baseline.*  *Proposal 14: ‘Tx and/or Rx Beam ID(s) of the N predicted DL Tx and/or Rx beams and other information’ can be considered with LOS probability.*  *Proposal 15: Benefits from utilization of TX/Rx beam angles should be clarified.* |
| Vivo[7] | *Proposal 11: Support to prioritize following AI output for further study on specification impact:*  * Tx and/or Rx Beam ID(s)/angle(s) and/or the predicted L1-RSRP of the N predicted DL Tx and/or Rx beams.*  * The N predicted Tx/Rx beams can be produced according to the expected beam information input to the AI model*  * FFS: study global beam ID or local beam ID*  * FFS: study global beam information, e.g. global beam ID or beam angle, with minimum exposures of implementation details*  *Proposal 12: Suggest to deprioritize Alt.2, i.e. Tx and/or Rx Beam ID(s) of the N predicted DL Tx and/or Rx beams and other information, for further study specification impact.* |
| Intel[9] | *Proposal 7: For BM-Case1 and 2, Alt-1 (Tx and/or Rx Beam ID(s) and/or the predicted L1-RSRP of the N predicted DL Tx and/or Rx beams) should be considered as the baseline use case, with potential specification impact on how beam IDs are mapped in the spatial domain.* |
| Sony[10] | *Proposal 1 : For the case of beam pair prediction, following output types of AI model at continuous dwelling time duration can be considered:*  *a) a DL Tx beam and a corresponding DL Rx beam*  *b) multiple DL Tx beams and a corresponding DL Rx beam*  *c) a DL Tx beam and multiple corresponding DL Rx beams*  *Proposal 6 : For the output of AI/ML, should clearly indicate the criterion associated with the predicted beam ID in BM-case1 and BM-case2, for example, sum probabilities of being the best beams higher than a threshold, maximum dwelling time, maximum RSRP, etc.* |
| Google[12] | *Proposal 3: For spatial domain beam prediction, support the best beam possibility for each beam in Set A as the output.*  *Proposal 4: For spatial-domain beam prediction, the output for Alt3 can be the channel eigenvector used for network beam generation.*  *Proposal 6: For spatial-domain beam prediction, study to predict the “weak” beam to facilitate the MU-MIMO UE pairing.*  *Proposal 10: For time-domain beam prediction, support the best beam possibility for each beam in Set A as the output.*  *Proposal 11: When AI/ML model is implemented in the NW side, the output for the AI/ML for time domain beam prediction with spec impact should be the reference angle for DL Rx beam refinement (Alt3).*  *Proposal 12: When AI/ML model is implemented in the UE side, the output for the AI/ML model for time domain beam prediction with spec impact should be the reference angle for DL Tx beam refinement (Alt3).* |
| NEC[15] | *Proposal 3: Support selecting Top-N DL Tx and/or Rx beams according to some pre-defined rules, e.g., a sum probability of being the best beam higher than a threshold, L1-RSRP higher than a threshold.* |
| Xiaomi[21] | *Proposal 6: Support Tx and/or Rx Beam ID(s) and/or the predicted L1-RSRP of the N predicted DL Tx and/or Rx beams as the AI/ML model output with high priority.* |
| OPPO[23] | *Proposal 24: For the output of AI/ML model for BM-Case1 and BM-Case2, suggest to include at least*  *• Tx and/or Rx Beam ID(s)*  *• The predicted L1-RSRP of the predicted Top-K DL Tx and/or Rx beams*  *• Note: the above output should be extended for F time instances for BM-Case2* |
| Lenovo[27] | *Proposal 5: When specifying the AI/ML model output, we should consider that it may be used for performance monitoring.* |
|  |  |
|  |  |

###### Mod’s assessment

**Mod’s assessment**: We will focus on the spec impact (if any) of AI model output in other section(s).

|  |  |
| --- | --- |
| Company | Comments |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

## Model/functionality selection, activation, deactivation, switching, and fallback operation

In previous RAN1 meeting(s), the following agreements were made:

|  |
| --- |
| **RAN1#110bis-e**  Agreement  For model selection, activation, deactivation, switching, and fallback at least for UE sided models and two-sided models, study the following mechanisms:   * Decision by the network   + Network-initiated   + UE-initiated, requested to the network * Decision by the UE   + Event-triggered as configured by the network, UE’s decision is reported to network   + UE-autonomous, UE’s decision is reported to the network   + UE-autonomous, UE’s decision is not reported to the network   FFS: for network sided models  FFS: other mechanisms  Agreement  Study the specification impact to support multiple AI models for the same functionality, at least including the following aspects:   * Procedure and assistance ignaling for the AI model switching and/or selection   FFS: Model selection refers to the selection of an AI/ML model among models for the same functionality. (Exact terminology to be discussed/defined)  Agreement (AI 9.2.1)  For model selection, activation, deactivation, switching, and fallback at least for UE sided models and two-sided models, study the following mechanisms:   * Decision by the network   + Network-initiated   + UE-initiated, requested to the network * Decision by the UE   + Event-triggered as configured by the network, UE’s decision is reported to network   + UE-autonomous, UE’s decision is reported to the network   + UE-autonomous, UE’s decision is not reported to the network   FFS: for network sided models  FFS: other mechanisms  **RAN1#113**  Agreement (AI 9.2.1)  For the purpose of activation/selection/switching of UE-side models/UE-part of two-sided models /functionalities (if applicable), study necessity, feasibility and potential specification impact for methods to assess/monitor the applicability and expected performance of an inactive model/functionality, including the following examples:   * Assessment/Monitoring based on the additional conditions associated with the model/functionality * Assessment/Monitoring based on input/output data distribution * Assessment/Monitoring using the inactive model/functionality for monitoring purpose and measuring the inference accuracy * Assessment/Monitoring based on past knowledge of the performance of the same model/functionality (e.g., based on other UEs)   FFS: Requirements for the assessment/monitoring to be reliable (e.g., sufficient data coverage during evaluation)  FFS: Additional aspects specific to the case where the inactive model has never been activated before, if any. |

The related proposals in tdocs are copied as below:

|  |  |
| --- | --- |
| NVIDIA[3] | *Proposal 6: For AI/ML based beam prediction in spatial/time domain, study potential specification impact related to assistance signalling and procedure for model configuration, model activation/deactivation, model recovery/termination, and model selection.* |
| ZTE[8] | *Proposal 29: The final decision on model/functionality selection/activation/deactivation/switching/fallback operation should be made by NW to guarantee overall NW performance.* |
| Sony[10] | *Proposal 9 : Model selection can be performed when there are changes in the application environment, which can be obtained through beam measurement results, e.g., L1-SINR, L1-RSRP etc.* |
| CATT[14] | *Proposal 16: Regarding the performance monitoring for BM-Case1 and BM-Case2, study the specification impacts on the following aspects:*  *• Model update/switching/fallback procedures based on model monitoring results, including the signaling exchange between the gNB and UE*  *• Trigger condition for model update/switching/fallback.* |
| Nokia[19] | *Proposal 7. For UE-sided BM-Case1 and BM-Case2, when the UE supports more than one functionality, the gNB shall be able to de-activate/switch one of the functionalities via dynamic signaling (e.g., MAC-CE).*  *Observation 11. For the purpose of switching/activating a UE-sided functionality, reporting the worse-case switching/activation delay for any of the supported functionalities is not sufficient as the actual delays requires for switching from one functionality to another functionality or activation/deactivation delay of a functionality is not considered.*  *Proposal 27. For UE-sided BM-Case1, RAN1 to consider reporting of actual delays required for switching from one functionality to another functionality or activation/deactivation delay of a functionality for the purpose of switching/activating a UE-sided functionality-based LCM.* |
| Lenovo[27] | *Proposal 14: Dynamic switching between AI/ML based beam prediction and non-AI/ML based beam report schemes as well as dynamic switching between different AI/ML models is supported.* |
|  |  |

###### Mod’s assessment

**Mod’s assessment**: Most proposals (if not all) are quite generic and lack sufficient details dedicated to BM sub use cases. Thus, they seem also applicable to other use case(s). Meanwhile, some discussions are related to or depends on other topics

**Mod’s suggestion**: Postpone the discussion to wait for more detailed proposal(s)

|  |  |
| --- | --- |
| Company | Comments |
|  |  |
|  |  |

## Misc

|  |  |
| --- | --- |
| FUTUREWEI[2] | *Proposal 7: Potential standard impacts related to supporting model generalization across scenarios and/or configurations needs to be further studied, for example, indication of a configuration change that may require additional pre-/post-processing or applying adaptation techniques.* |
| IDC[6] | *Observation 18: The current NR specification does not consider association between beams with different beam widths.*  *Observation 19: Utilizing association between beams with different beam widths can provide benefits for prediction accuracy e.g., robust estimation/identification of whole spatial characteristics with wide beams and accurate beam identification with narrow beams.*  *Proposal 32: Study benefits of specification enhancements on association between beams with different beam widths.* |
| Vivo[7] | *Proposal 15: For case 1 and case 2 of beam management, both collaboration level-y, and collaboration level-z can be considered.*  *Proposal 16: Take the following supportable model update choices as one aspect for defining model update levels of beam management.*  *- Choice 0: No model update during lifecycle management*  *- Choice 1: Updating model parameters w/o model transfer*  *- Choice 2: Updating model parameters with model transfer*  *- Study the lifecycle management signaling and procedures for each of the collaboration levels and model updating choices.* |
| Ericsson[11] | *Proposal 1 Include AI/ML based beam management in Rel-19 work item. Based on the few agreements made for specification impact discussion, focus on a limited set of sub-use cases and alternatives in each use case (e.g., possible model input/output).*  *Proposal 2 Support both UE-side model and gNB-side model in Rel-19 work item, where training and inference is done in the same node.*  *Proposal 3 Consider spatial TX beam prediction for the Rel-19 work item.*  *Proposal 4 Focus on UE L1-RSRP measurement as input data for a Rel-19 work item.*  *Proposal 5 Keep all AI PHY use cases that are mature enough for normative work in one single work item in Rel-19.* |
| Apple[20] | *Observation 1: AI/ML models can be crafted as a universal channel parameter estimator with good generalization or as a beam management database and associated query mechanism customized for a specific cell which is not expected to generalize well.* |
| Xiaomi[21] | *Proposal 5: Consider a common AI model for UE with different number of Rx beam.* |
| OPPO[23] | *Proposal 19: For BM-Case1 and BM-Case2, consider the UE capability on AI/ML beam prediction at later stage.*  *Proposal 25: For BM-Case1 and BM-Case2, study enhancement on generalization of AI/ML model (if necessary) under heterogeneous scenarios and different Tx and/or Rx beam configurations.* |
| Lenovo[27] | *Proposal 6: Introduce AI/ML processing unit concept for high efficiency AI/ML resource management.* |
|  |  |

|  |  |
| --- | --- |
| Company | Comments |
|  |  |
|  |  |

# Summary of Discussion

## Wednesday 1st online session

**15 companies proposed mechanism for overhead reduction**

***Proposal 2.2.2: Regarding data collection for NW-side AI/ML model of BM-Case1 and BM-Case2, the following approaches have been studied by companies for overhead reduction***

* ***the omission/selection of collected data***
* ***the compression of collected data***
* ***Note1: For the different purposes of data collection, the overhead reduction mechanisms and corresponding specification impacts may be different.***
* ***Note2: Support of any mechanism(s) (if necessary) for each LCM purpose and the potential spec impact (if any) are separate discussions***
* ***Note 3: UE complexity and power consumption should be considered.***

***Proposal 4.3.1:******For BM-Case1 and BM-Case2 with a UE-side AI/ML model, regarding UE-side performance monitoring for a functionality, specification impact is identified to support NW to indicate UE to do LCM operations at functionality level.***

***Proposal 3.2.1:***

***Conclusion***

***For BM-Case1 and BM-Case2 with a UE-side AI/ML model, for AI model inference, the legacy mechanism can be used to perform beam indication of beams in Set A not in Set B***

* ***No consensus on the need of any specification enhancement from RAN1 perspective***

Summary of companies’ view

* 14 companies support
* 4 companies not support (LGE, Sony, Apple, QC )
  1. Wednesday 2nd online session

***Proposal 2.2.1: Regarding data collection for NW-side AI/ML model of BM-Case1 and BM-Case2, the following reporting signaling for beam-specific aspects are considered:***

* ***L1 signaling to report the collected data***
* ***Higher-layer signaling to report the collected data*** 
  + ***At least not applicable to AI/ML model inference***
* ***Note1: higher layer signaling design is up to other WG(s)***
* ***Note2: Whether each signaling applicable to each ~~LMC~~ LCM purpose is a separate discussion***
* ***Note3: The legacy signaling principle (e.g. RSRP reporting for L1) can be re-used***

## Thursday online session

***Proposal 4.1.1:***

***Observation:***

***Regarding the performance metric(s) of AI/ML model monitoring for BM-Case1 and BM-Case2, the following table is identified***

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Alt.1: Beam prediction accuracy related KPIs, e.g., Top-K/1 beam prediction accuracy | Alt.2: Link quality related KPIs, e.g., throughput, L1-RSRP, L1-SINR, hypothetical BLER | Alt.3: Performance metric based on input/output data distribution of AI/ML | Alt.4: The L1-RSRP difference evaluated by comparing measured RSRP and predicted RSRP |
| Applicable to different AI models | Applicable to all AI models | Applicable to all AI models | May not applicable to some implementation of AI model (e.g., not output of predicted L1-RSRP) | May not applicable to some implementation of AI model (e.g., not output of predicted L1-RSRP) |
| Performance indication | Reflect the prediction accuracy of AI model | Reflect the system/link performance | Reflect the change of the statics of the input/output data | Reflect accuracy of the predicted 1-RSRP |
| Other aspects | Not reflect the system/link performance directly | Not reflect the prediction accuracy of AI model directly | Not reflect the prediction performance of AI model directly  Not reflect the system/link performance directly | Not reflect the system/link performance directly |

Note1: The above analysis shall not give an indication about whether/which metric is supported or specified

Note2: Monitoring performance of the above alternatives are not touched in the table

# Reference

1. R1-2306399 Discussion on other aspects of AI/ML beam management New H3C Technologies Co., Ltd.
2. R1-2306434 Discussion on other aspects of AI/ML for beam management FUTUREWEI
3. R1-2306478 AI and ML for beam management NVIDIA
4. R1-2306514 Discussion on AI/ML for beam management Huawei, HiSilicon
5. R1-2306640 Other aspects on AI/ML for beam management Spreadtrum Communications
6. R1-2306690 Discussion for other aspects on AI/ML for beam management InterDigital, Inc.
7. R1-2306743 Other aspects on AI/ML for beam management vivo
8. R1-2306798 Discussion on other aspects for AI beam management ZTE
9. R1-2306857 Other Aspects on AI/ML for Beam Management Intel Corporation
10. R1-2306904 Considerations on AI/ML for beam management Sony
11. R1-2306929 Discussion on AI/ML for beam management Ericsson
12. R1-2306960 On Enhancement of AI/ML based Beam Management Google
13. R1-2307016 Other aspects on AI/ML for beam management LG Electronics
14. R1-2307079 Discussion on other aspects for AI/ML beam management CATT
15. R1-2307137 Discussion on AI ML for beam management NEC
16. R1-2307157 Discussion for specification impacts on AI/ML for beam management Fujitsu
17. R1-2307186 Discussion on other aspects on AI/ML for beam management CMCC
18. R1-2307233 Discussion on AI/ML for beam management Panasonic
19. R1-2307241 Other aspects on ML for beam management Nokia, Nokia Shanghai Bell
20. R1-2307271 Discussion on other aspects of AI/ML based beam management enhancements Apple
21. R1-2307378 Potential specification impact on AI/ML for beam management xiaomi
22. R1-2307469 Discussion on other aspects on AI/ML for beam management NTT DOCOMO, INC.
23. R1-2307567 Other aspects of AI/ML for beam management OPPO
24. R1-2307671 Discussion on potential specification impact for beam management Samsung
25. R1-2307730 Prediction of untransmitted beams in a UE-side AI-ML model Rakuten Symphony
26. R1-2307742 Discussion on other aspects on AI/ML for beam management ETRI
27. R1-2307809 Further aspects of AI/ML for beam management Lenovo
28. R1-2307863 Discussions on AI-ML for Beam management CAICT
29. R1-2307867 Discussion on other aspects on AI/ML for beam management KT Corp.
30. R1-2307919 Other aspects on AI/ML for beam management Qualcomm Incorporated
31. R1-2308055 Other aspects on AI/ML for beam management MediaTek Inc.
32. R1-2308160 Discussion on other aspects of AI/ML for beam management Indian Institute of Technology Madras (IITM), IIT Kanpur, CEWiT

# Appendix A: Contact Information

The following information was collected in the last meeting(s). Please feel free to add/update/correct contact information if needed.

|  |  |  |
| --- | --- | --- |
| Company | Name | Email |
| Moderator | Zhihua SHI | szh@oppo.com |
| Apple | Weidong Yang | Wyang23@apple.com |
| vivo | Peng SUN | sunpeng@vivo.com |
| AT&T | Thomas Novlan | thomas\_novlan@labs.att.com |
| Futurewei | Chunhui Zhu | czhu@futurewei.com |
| Xiaomi | Mingju Li | limingju@xiaomi.com |
| Lenovo | Bingchao LIU | liubc2@lenovo.com |
| Sony | Chen SUN | Chen.sun@sony.com |
| Huawei, HiSilicon | Thorsten Schier | thorsten.schier@huawei.com |
| NEC | Zhen He | he\_zhen@nec.cn |
| LG Electronics | Jiwon Kang  Haewook Park | [jw.kang@lge.com](mailto:jw.kang@lge.com)  [haewook.park@lge.com](mailto:haewook.park@lge.com) |
| Panasonic | Quan Kuang | quan.kuang@eu.panasonic.com |
| Ericsson | Henrik Ryden | Henrik.a.ryden@ericsson.com |
| Nokia, NSB | Keeth Jayasinghe  Mihai Enescu | keeth.jayasinghe@nokia.com, mihai.enescu@nokia.com |
| CATT | Yongqiang FEI | feiyongqiang@catt.cn |
| Fujitsu | Wang Xin | wangxin@fujitsu.com |
| Samsung | Zhe Chen | tom.chenzhe@samsung.com |
| CMCC | Jiazhen Zhang | zhangjiazhen@chinamobile.com |
| NVIDIA | Xingqin Lin | xingqinl@nvidia.com |
| CAICT | Xiaofeng Liu | Liuxiaofeng1@caict.ac.cn |
| OPPO | Jianfei Cao | caojianfei@oppo.com |
| MediaTek | Gyu Bum Kyung  Yu-Jen Ku | gyubum.kyung@mediatek.com  yu-jen.ku@mediatek.com |
| Intel | Avik Sengupta | avik.sengupta@intel.com |
| NTT DOCOMO | Haruhi Echigo | haruhi.echigo.fw@nttdocomo.com |
| Beijing Jiaotong University (BJTU) | Wei Chen | weich@bjtu.edu.cn |
| ZTE, Sanechips | Wenfeng Liu | liu.wenfeng@zte.com.cn |
| InterDigital | Youngwoo Kwak | youngwoo.kwak@interdigital.com |
| Qualcomm | Hamed Pezeshki | hamedp@qti.qualcomm.com |
| Spreadtrum | Dawei Ma | dawei.ma@unisoc.com |
| Charter Communications | Dumitru M. Ionescu  Samer Henry | dumitru.ionescu@charter.com  C-Samer.Henry@charter.com |
|  |  |  |

# Appendix B: Agreements

## RAN#114

## RAN1#113

Agreement

For BM-Case1 and BM-Case2 with a UE-side AI/ML model, regarding performance monitoring, study potential spec impact(s) from the following aspects in addition to those included in previous agreements:

* Configuration/Signalling from gNB to UE for measurement and/or reporting
* UE calculates performance metric(s), either reports it to NW or reports an event to NW based on the performance metric(s)
  + FFS: definition of an event and the performance metric(s) used to identify it
* Indication from NW for UE to do LCM operations

Agreement

For BM-Case1 and BM-Case2 with a UE-side AI/ML model, regarding performance monitoring, study the necessity and potential spec impact(s) of the mechanism that facilitate UE to detect whether the functionality/model is suitable or no longer suitable.

Conclusion

For the study of DL beam pair prediction of BM-Case1 and BM-Case2 with a UE-side AI/ML model, RAN1 has no consensus to support the reporting of the predicted Rx beam(s) (e.g., Rx beam ID, Rx beam angle information, etc) from UE to network.

Agreement

For BM-Case2, study necessity, benefit(s) and potential specification impact from the following additional aspects for AI model inference:

* Reporting information about measurements of multiple past time instances in one reporting instance for BM-Case2
  + Note: only applicable to network-side AI/ML model
* Note: The potential performance gains of measurement reporting should be justified by considering UCI payload overhead

Agreement

For BM-Case1 and BM-Case2, study necessity, benefit(s) and potential specification impact from the following additional aspects for AI model inference:

* How to perform beam indication of beams in Set A not in Set B
  + Note: the legacy mechanism may be sufficient

Agreement

Regarding data collection for BM-Case1 and BM-Case2 with a UE-side AI/ML model, study the benefits, necessity and potential specification impact of the following aspect on top of those we have agreed in previous meeting:

* Assistance information from NW to UE for UE data collection for categorizing the data for the purpose of differentiating characteristics of data
  + The assistance information should preserve privacy/proprietary information.

Agreement

For BM-Case1 and BM-Case2 with a UE-side AI/ML model, study the necessity and potential BM-specific conditions/additional conditions for functionality(ies) and/or model(s) at least from the following aspects:

* information regarding model inference
* Set A / Set B configuration
* performance monitoring
* data collection
* assistance information

## RAN1#112bis-e

Agreement

Regarding the data collection at UE side for UE-side AI/ML model, study the potential specification impact of UE reporting to network from the following aspect

* Supported/preferred configurations of DL RS transmission
* Other aspect(s) is not precluded

Agreement

Regarding the data collection at UE side for UE-side AI/ML model, study the potential specification impact (if any) to initiate/trigger data collection from RAN1 point of view by considering the following options as a starting point

* Option 1: data collection initiated/triggered by configuration from NW
* Option 2: request from UE for data collection
  + FFS: details

Agreement

Regarding data collection for NW-side AI/ML model, study the following options (including the combination of options) for the contents of collected data,

* Opt.1: M1 L1-RSRPs (corresponding to M1 beams) with the indication of beams (beam pairs) based on the measurement corresponding to a beam set, where M1 can be larger than 4, if applicable
  + FFS: the range of M1
* Opt.2: M2 L1-RSRPs (corresponding to M2 beams) based on the measurement corresponding to a beam set, where M2 can be larger than 4, if applicable
  + FFS: the range of M2
* Opt.3: M3 beam (beam pair) indices based on the measurement corresponding to a beam set, where M3 can be larger than 4, if applicable
  + FFS: the range of M3
* FFS: How to select the M1/M2/M3 beam(s) or beam pair(s)
* Note: Overhead, UE complexity and power consumption should be considered for the above options

Agreement

Regarding data collection for NW-side AI/ML model, study necessity, benefits and beam-management-specific potential specification impact from RAN1 point of view on the following additional aspects

* Mechanism related to the reporting
* Additional information for content of the reporting
  + FFS: Information associated with or configured for the reported data samples, e.g., timestamps, SNR, data quality, etc.
* Reporting overhead reduction
* Note1: non-3GPP based solution is a separate issue.
* Note2: The framework corresponding to higher layer(s) are up to the associated WG(s)
* Note 3: Overhead, UE complexity and power consumption should be considered

Agreement

For AI/ML performance monitoring for BM-Case1 and BM-Case2, study potential specification impact of at least the following alternatives as the benchmark/reference (if applicable) for performance comparison:

* Alt.1: The best beam(s) obtained by measuring beams of a set indicated by gNB (e.g., Beams from Set A)
  + FFS: gNB configures one or multiple sets for one or multiple benchmarks/references
* Alt.4: Measurements of the predicted best beam(s) corresponding to model output (e.g., Comparison between actual L1-RSRP and predicted RSRP of predicted Top-1/K Beams)
* FFS:
  + Alt.3: The beam corresponding to some or all the indicated/activated TCI state(s)
* Other alternative is not precluded.

## RAN1#112

Conclusion

For the sub use case BM-Case1 and BM-Case2, “Alt.2: DL Rx beam prediction” is deprioritized.

Agreement

For BM-Case1 and BM-Case2 with a UE-side AI/ML model, study the necessity, feasibility and the potential specification impact (if needed) of the following information reported from UE to network:

* Predicted L1-RSRP(s) corresponding to the DL Tx beam(s) or beam pair(s)
  + Whether/how to differentiate predicted L1-RSRP and measured L1-RSRP
* Confidence/probability information related to the output of AI/ML model inference (e.g., predicted beams)
  + FFS: Definition/content of confidence/probability information
* Note: At least the performance and spec impact should be considered

Agreement

For BM-Case1 and BM-Case2 with a UE-side AI/ML model, study potential specification impact of AI model inference from the following additional aspects on top of previous agreements:

* Indication of the associated Set A from network to UE, e.g., association/mapping of beams within Set A and beams within Set B if applicable
* Beam indication from network for UE reception
* Note: The second bullet may or may not have additional specification impact (e.g., legacy mechanism may be reused).

Conclusion

Regarding the explicit assistance information from UE to network for NW-side AI/ML model, RAN1 has no consensus to support the following information

* UE location
* UE moving direction
* UE Rx beam shape/direction

Conclusion

Regarding the explicit assistance information from network to UE for UE-side AI/ML model, RAN1 has no consensus to support the following information

* NW-side beam shape information
  + E.g., 3dB beamwidth, beam boresight directions, beam shape, Tx beam angle, etc.
* Note: Other information (e.g., relative information) of Tx beam(s) preserving sensitive proprietary information is a separate discussion
  + e.g., some information following the same principle of Rel-17 positioning agreement

Agreement

Regarding the performance metric(s) of AI/ML model monitoring for BM-Case1 and BM-Case2, study the following alternatives (including feasibility/necessity) with potential down-selection:

* Alt.1: Beam prediction accuracy related KPIs, e.g., Top-K/1 beam prediction accuracy
* Alt.2: Link quality related KPIs, e.g., throughput, L1-RSRP, L1-SINR, hypothetical BLER
* Alt.3: Performance metric based on input/output data distribution of AI/ML
* Alt.4: The L1-RSRP difference evaluated by comparing measured RSRP and predicted RSRP
* Other alternatives are not precluded
* Note: At least the performance and spec impact should be considered

Agreement

For BM-Case1 and BM-Case2 with a UE-side AI/ML model, regarding NW-side performance monitoring, study the following aspects as a starting point including the study of necessity:

* Configuration/Signaling from gNB to UE for measurement and/or reporting
* UE reporting to NW (e.g., for the calculation of performance metric)
* Indication from NW for UE to do LCM operations
* Other aspect(s) is not precluded
* Note1: At least the performance and reporting overhead of model monitoring mechanism should be considered

Agreement

For BM-Case1 and BM-Case2 with a UE-side AI/ML model, regarding UE-side performance monitoring, study the following aspects as a starting point including the study of necessity and feasibility:

* Indication/request/report from UE to gNB for performance monitoring
  + Note: The indication/request/report may be not needed in some case(s)
* Configuration/Signaling from gNB to UE for performance monitoring
* Other aspect(s) is not precluded

## RAN1#111

Agreement

For the sub use case BM-Case1 and BM-Case2, at least support Alt.1 and Alt.2 for AI/ML model training and inference for further study:

* Alt.1. AI/ML model training and inference at NW side
* Alt.2. AI/ML model training and inference at UE side
* The discussion on Alt.3 for BM-Case1 and BM-Case2 is dependent on the conclusion/agreement of Agenda item 9.2.1 of RAN1 and/or RAN2 on whether to support model transfer for UE-side AI/ML model or not
  + Alt.3. AI/ML model training at NW side, AI/ML model inference at UE side

Agreement

Regarding the data collection for AI/ML model training at UE side, study the potential specification impact considering the following additional aspects.

* Whether and how to initiate data collection
* Configurations, e.g., configuration related to set A and/or Set B, information on association/mapping of Set A and Set B
* Assistance information from Network to UE (If supported)
* Other aspect(s) is not precluded

Agreement

For BM-Case1 and BM-Case2 with a network-side AI/ML model, study potential specification impact on the following L1 reporting enhancement for AI/ML model inference

* UE to report the measurement results of more than 4 beams in one reporting instance
* Other L1 reporting enhancements can be considered

Agreement

Regarding NW-side model monitoring for a network-side AI/ML model of BM-Case1 and BM-Case2, study the necessity and the potential specification impacts from the following aspects:

* UE reporting of beam measurement(s) based on a set of beams indicated by gNB.
* Signaling, e.g., RRC-based, L1-based.
* Note: Performance and UE complexity, power consumption should be considered.

## RAN1#110bis-e

Conclusion

For AI/ML based beam management, RAN1 has no consensus to support on studying any other sub use case in addition to BM-Case1 and BM-Case2

Note: this conclusion is independent of the discussion on the alternatives of AI/ML model inputs for BM-Case1 and BM-Case2

Conclusion

For the sub use case BM-Case1 and BM-Case2, Set B is a set of beams whose measurements are taken as inputs of the AI/ML model

***Agreement***

***For BM-Case1 with a UE-side AI/ML model, study the potential specification impact of L1 signaling to report the following information of AI/ML model inference to NW***

* ***The beam(s) that is based on the output of AI/ML model inference***
* ***FFS: Predicted L1-RSRP corresponding to the beam(s)***
* ***FFS: other information***

***Agreement***

***For BM-Case2 with a UE-side AI/ML model, study the potential specification impact of L1 signaling to report the following information of AI/ML model inference to NW***

* ***The beam(s)*** ***of N future time instance(s) that is based on the output of AI/ML model inference***
  + ***FFS: value of N***
* ***FFS: Predicted L1-RSRP corresponding to the beam(s)***
* ***Information about the timestamp corresponding the reported beam(s)***
  + ***FFS: explicit or implicit***
* ***FFS: other information***

***Working Assumption***

***For BM-Case1 and BM-Case2 with a network-side AI/ML model, study the following L1 beam reporting enhancement for AI/ML model inference***

* ***UE to report the measurement results of more than 4 beams in one reporting instance***
* ***Other L1 reporting enhancements can be considered***

***Agreement***

***For BM-Case1 and BM-Case2 with a network-side AI/ML model, study the NW-side model monitoring:***

* ***NW monitors the performance metric(s) and makes decision(s) of model selection/activation/ deactivation/switching/ fallback operation***

***Agreement***

***Regarding NW-side model monitoring for a network-side AI/ML model of BM-Case1 and BM-Case2, study the potential specification impacts from the following aspects***

* ***Beam measurement and report for model monitoring***
* ***Note: This may or may not have specification impact.***

***Agreement***

***For BM-Case1 and BM-Case2 with a UE-side AI/ML model, study the following alternatives for model monitoring with potential down-selection:***

* ***Atl1. UE-side Model monitoring***
  + ***UE monitors the performance metric(s)***
  + ***UE makes decision(s) of model selection/activation/ deactivation/switching/fallback operation***
* ***Atl2. NW-side Model monitoring***
  + ***NW monitors the performance metric(s)***
  + ***NW makes decision(s) of model selection/activation/ deactivation/switching/ fallback operation***
* ***Alt3. Hybrid model monitoring***
  + ***UE monitors the performance metric(s)***
  + ***NW makes decision(s) of model selection/activation/ deactivation/switching/ fallback operation***

## RAN1#110

Agreement

For the sub use case BM-Case1, support the following alternatives for further study:

* Alt.1: Set A and Set B are different (Set B is NOT a subset of Set A)
* Alt.2: Set B is a subset of Set A
* Note1: Set A is for DL beam prediction and Set B is for DL beam measurement.
* Note2: The beam patterns of Set A and Set B can be clarified by the companies.

Agreement

For the sub use case BM-Case2, further study the following alternatives:

* Alt.1: Set A and Set B are different (Set B is NOT a subset of Set A)
* Alt.2: Set B is a subset of Set A (Set A and Set B are not the same)
* Alt.3: Set A and Set B are the same
* Note1: The beam pattern of Set A and Set B can be clarified by the companies.

Agreement

For the data collection for AI/ML model training (if supported), study the following aspects as a starting point for potential necessary specification impact:

* Signaling/configuration/measurement/report for data collection, e.g., signaling aspects related to assistance information (if supported), Reference signals
* Content/type of the collected data
* Other aspect(s) is not precluded

Agreement

At least for the sub use case BM-Case1 and BM-Case2, support both Alt.1 and Alt.2 for the study of AI/ML model training:

* Alt.1: AI/ML model training at NW side;
* Alt.2: AI/ML model training at UE side.

Note: Whether it is online or offline training is a separate discussion.

Agreement

For the sub use case BM-Case1 and BM-Case2, further study the following alternatives for the predicted beams:

* Alt.1: DL Tx beam prediction
* Alt.2: DL Rx beam prediction
* Alt.3: Beam pair prediction (a beam pair consists of a DL Tx beam and a corresponding DL Rx beam)
* Note1: DL Rx beam prediction may or may not have spec impact

Agreement

Regarding the model monitoring for BM-Case1 and BM-Case2, to investigate specification impacts from the following aspects

* Performance metric(s)
* Benchmark/reference for the performance comparison
* Signaling/configuration/measurement/report for model monitoring, e.g., signaling aspects related to assistance information (if supported), Reference signals
* Other aspect(s) is not precluded

Agreement

In order to facilitate the AI/ML model inference, study the following aspects as a starting point:

* Enhanced or new configurations/UE reporting/UE measurement, e.g., Enhanced or new beam measurement and/or beam reporting
* Enhanced or new signaling for measurement configuration/triggering
* Signaling of assistance information (if applicable)
* Other aspect(s) is not precluded

Agreement

Regarding the sub use case BM-Case1 and BM-Case2, study the following alternatives for AI/ML output:

* Alt.1: Tx and/or Rx Beam ID(s) and/or the predicted L1-RSRP of the N predicted DL Tx and/or Rx beams
  + E.g., N predicted beams can be the top-N predicted beams
* Alt.2: Tx and/or Rx Beam ID(s) of the N predicted DL Tx and/or Rx beams and other information
  + FFS: other information (e.g., probability for the beam to be the best beam, the associated confidence, beam application time/dwelling time, Predicted Beam failure)
  + E.g., N predicted beams can be the top-N predicted beams
* Alt.3: Tx and/or Rx Beam angle(s) and/or the predicted L1-RSRP of the N predicted DL Tx and/or Rx beams
  + E.g., N predicted beams can be the top-N predicted beams
  + FFS: details of Beam angle(s)
* FFS: how to select the N DL Tx and/or Rx beams (e.g., L1-RSRP higher than a threshold, a sum probability of being the best beams higher than a threshold, RSRP corresponding to the expected Tx and/or Rx beam direction(s))
* Note1: It is up to companies to provide other alternative(s)
* Note2: Beam ID is only used for discussion purpose
* Note3: All the outputs are “nominal” and only for discussion purpose
* Note4: Values of N is up to each company.
* Note5: All of the outputs in the above alternatives may vary based on whether the AI/ML model inference is at UE side or gNB side.
* Note 6: The Top-N beam IDs might have been derived via post-processing of the ML-model output

## RAN1#109-e

Agreement

For AI/ML-based beam management, support BM-Case1 and BM-Case2 for characterization and baseline performance evaluations

* BM-Case1: Spatial-domain DL beam prediction for Set A of beams based on measurement results of Set B of beams
* BM-Case2: Temporal DL beam prediction for Set A of beams based on the historic measurement results of Set B of beams
* FFS: details of BM-Case1 and BM-Case2
* FFS: other sub use cases

Note: For BM-Case1 and BM-Case2, Beams in Set A and Set B can be in the same Frequency Range

Agreement

Regarding the sub use case BM-Case2, the measurement results of K (K>=1) latest measurement instances are used for AI/ML model input:

* The value of K is up to companies

Agreement

Regarding the sub use case BM-Case2, AI/ML model output should be F predictions for F future time instances, where each prediction is for each time instance.

* At least F = 1
* The other value(s) of F is up to companies

Agreement

For the sub use case BM-Case1, consider both Alt.1 and Alt.2 for further study:

* Alt.1: AI/ML inference at NW side
* Alt.2: AI/ML inference at UE side

Agreement

For the sub use case BM-Case2, consider both Alt.1 and Alt.2 for further study:

* Alt.1: AI/ML inference at NW side
* Alt.2: AI/ML inference at UE side

Conclusion

For the sub use case BM-Case1, consider the following alternatives for further study:

* Alt.1: Set B is a subset of Set A
  + FFS: the number of beams in Set A and B
  + FFS: how to determine Set B out of the beams in Set A (e.g., fixed pattern, random pattern, …)
* Alt.2: Set A and Set B are different (e.g. Set A consists of narrow beams and Set B consists of wide beams)
  + FFS: the number of beams in Set A and B
  + FFS: QCL relation between beams in Set A and beams in Set B
  + ~~FFS: construction of Set B (e.g., regular pre-defined codebook, codebook other than regular pre-defined one)~~
* Note1: Set A is for DL beam prediction and Set B is for DL beam measurement.
* Note2: The narrow and wide beam terminology is for SI discussion only and have no specification impact
* Note3: The codebook constructions of Set A and Set B can be clarified by the companies.

Conclusion

For the sub use case BM-Case2, further study the following alternatives with potential down-selection:

* Alt.1: Set A and Set B are different (e.g. Set A consists of narrow beams and Set B consists of wide beams)
  + FFS: QCL relation between beams in Set A and beams in Set B
* Alt.2: Set B is a subset of Set A (Set A and Set B are not the same)
  + FFS: how to determine Set B out of the beams in Set A (e.g., fixed pattern, random pattern, …)
* Alt.3: Set A and Set B are the same
* Note1: Predicted beam(s) are selected from Set A and measured beams used as input are selected from Set B.
* Note2: It is up to companies to provide other alternative(s)
* Note3: The narrow and wide beam terminology is for SI discussion only and have no specification impact

Conclusion

Regarding the sub use case BM-Case1, further study the following alternatives for AI/ML input:

* Alt.1: Only L1-RSRP measurement based on Set B
* Alt.2: L1-RSRP measurement based on Set B and assistance information
  + FFS: Assistance information. The following were mentioned by companions in the discussion:  Tx and/or Rx beam shape information (e.g., Tx and/or Rx beam pattern, Tx and/or Rx beam boresight direction (azimuth and elevation), 3dB beamwidth, etc.), expected Tx and/or Rx beam for the prediction (e.g., expected Tx and/or Rx angle, Tx and/or Rx beam ID for the prediction), UE position information, UE direction information, Tx beam usage information, UE orientation information, etc.
    - Note: The provision of assistance information may be infeasible due to the concern of disclosing proprietary information to the other side.
* Alt.3: CIR based on Set B
* Alt.4: L1-RSRP measurement based on Set B and the corresponding DL Tx and/or Rx beam ID
* Note1: It is up to companies to provide other alternative(s) including the combination of some alternatives
* Note2: All the inputs are “nominal” and only for discussion purpose.

Conclusion

Regarding the sub use case BM-Case2, further study the following alternatives of measurement results for AI/ML input (for each past measurement instance):

* Alt.1: Only L1-RSRP measurement based on Set B
* Alt 2: L1-RSRP measurement based on Set B and assistance information
  + FFS: Assistance information. The following were mentioned by companies in the discussion:, Tx and/or Rx beam angle, position information, UE direction information, positioning-related measurement (such as Multi-RTT), expected Tx and/or Rx beam/occasion for the prediction (e.g., expected Tx and/or Rx beam angle for the prediction, expected occasions of the prediction), Tx and/or Rx beam shape information (e.g., Tx and/or Rx beam pattern, Tx and/or Rx beam boresight directions (azimuth and elevation), 3dB beamwidth, etc.) , increase ratio of L1-RSRP for best N beams, UE orientation information
    - Note: The provision of assistance information may be infeasible due to the concern of disclosing proprietary information to the other side.
* Alt.3: L1-RSRP measurement based on Set B and the corresponding DL Tx and/or Rx beam ID
* Note1: It is up to companies to provide other alternative(s) including the combination of some alternatives
* Note2: All the inputs are “nominal” and only for discussion purpose.