**3GPP TSG RAN WG1#118 R1-2407322**

**Maastricht, NL, August 19th – 23rd, 2024**

**Agenda item:** **9.1.1**

**Source: Samsung (Moderator)**

**Title: FL summary #2 for AI/ML in beam management**

**Document for: Discussion and Decision**

# Introduction

In RAN#102, Rel-19 work item on “New WID on Artificial Intelligence (AI)/Machine Learning (ML) for NR Air Interface” is endorsed. The objective of the work item is as follows.

|  |
| --- |
| Provide specification support for the following aspects:   * Beam management - DL Tx beam prediction for both UE-sided model and NW-sided model, encompassing [RAN1/RAN2]:   + Spatial-domain DL Tx beam prediction for Set A of beams based on measurement results of Set B of beams (“BM-Case1”)   + Temporal DL Tx beam prediction for Set A of beams based on the historic measurement results of Set B of beams (“BM-Case2”)   + Specify necessary signalling/mechanism(s) to facilitate LCM operations specific to the Beam Management use cases, if any   + Enabling method(s) to ensure consistency between training and inference regarding NW-side additional conditions (if identified) for inference at UE   NOTE: Strive for common framework design to support both BM-Case1 and BM-Case2 |

In this contribution, summarized the contributions in RAN 1 #118 on AI/ML for beam management.

## (FL0) Question 0

* Please consider entering contact info below for the points of contact for this email discussion.

|  |  |  |
| --- | --- | --- |
| Company | Contact | Email address |
| New H3C | Lei Zhou | Zhou.leih@h3c.com |
| Indian Institute of Technology Madras | Sai Prasad Pirati | venkatasiva@5gtbiitm.in |
| OPPO | Jeffrey (Jianfei) Cao | [caojianfei@oppo.com](mailto:caojianfei@oppo.com) |
| Ericsson | Henrik Ryden | Henrik.a.ryden@ericsson.com |
| Fujitsu | WANG Guotong (David) | wangguotong@fujitsu.com |
| MediaTek | Yu-Jen Ku | [yu-jen.ku@mediatek.com](mailto:yu-jen.ku@mediatek.com) |
| TCL | Pu Yuan | pu.yuan@tcl.com |
| CMCC | Yi Zheng  Jiazhen Zhang | zhengyi@chinamobile.com  zhangjiazhen@chinamobile.com |
| NTT DOCOMO | Haruhi Echigo | haruhi.echigo.fw@nttdocomo.com |
| InterDigital | Youngwoo Kwak | [youngwoo.kwak@interdigital.com](mailto:youngwoo.kwak@interdigital.com) |
| vivo | Hao Wu | hao.wu@vivo.com |
| NEC | Peng Guan  Pravjyot Deogun  Zhen He  Yi Jiang | [Guan\_peng@nec.cn](mailto:Guan_peng@nec.cn)  [pravjyot.deogun@EMEA.NEC.COM](mailto:pravjyot.deogun@EMEA.NEC.COM)  he\_zhen@nec.cn  y-jiang\_ct@nec.com |
| ETRI | Yongjin Kwon | yjkwon@etri.re.kr |
| Xiaomi | Mingju | limingju@xiaomi.com |
| ZTE | Wenfeng LIU  Xingguang WEI | [liu.wenfeng@zte.com.cn](mailto:Liu.wenfeng@zte.com.cn)  wei.xingguang@zte.com.cn |
| Qualcomm | Hamed Pezeshki | hamedp@qti.qualcomm.com |
| Spreadtrum | Shijia Shao | [Shijia.shao@unisoc.com](mailto:Shijia.shao@unisoc.com)  Hualei.wang@unisoc.com |
| Panasonic | Henry Tran | [xuantuong.tran@sg.panasonic.com](mailto:xuantuong.tran@sg.panasonic.com) |
| CATT | Min Zhu | zhumin@catt.cn | |
| CEWiT | Ebin Chacko  Shiv Shankar | [echacko@cewit.org.in](mailto:echacko@cewit.org.in)  [shiv@cewit.org.in](mailto:shiv@cewit.org.in) | |
| Google | Yushu Zhang | yushuzhang@google.com | |
| Sharp | Liqing Liu | [liu.liqing@sharp.co.jp](mailto:liu.liqing@sharp.co.jp) | |
| Futurewei | Zhigang Rong | [zrong@futurewei.com](mailto:zrong@futurewei.com) | |
| LG | Seongwon Go | sw.go@lge.com | |
| Lenovo | Bingchao Liu | [Liubc2@lenovo.com](mailto:Liubc2@lenovo.com) | |
| Fraunhofer HHI | Baris Göktepe | Baris.goektepe@hhi.fraunhofer.de | |
| KDDI | Taishi Watanabe | ta-watanabe@kddi.com | |
| NVIDIA | Xingqin Lin | [xingqinl@nvidia.com](mailto:xingqinl@nvidia.com) | |
| SONY | Chen Sun  Yingshuang Bai | [chen.sun@sony.com](mailto:chen.sun@sony.com)  [yingshuang.bai@sony.com](mailto:yingshuang.bai@sony.com) | |
| Hw/HiSi | Thorsten Schier | [Thorsten.schier@huawei.com](mailto:Thorsten.schier@huawei.com) | |
| Intel | Debdeep Chatterjee | debdeep.chatterjee@intel.com | |
| Apple | Weidong Yang | wyang23@apple.com | |
| CAICT | Xiaofeng Liu | Liuxiaofeng1@caict.ac.cn | |
| China Telecom | Bei Yang  Wei Li | [yangbei1@chinatelecom.cn](mailto:yangbei1@chinatelecom.cn)  [liw40@chinatelecom.cn](mailto:liw40@chinatelecom.cn) | |
| Kyocera | Omar Sleem  Amit Kalhan | [omar.sleem@kyocera.com](mailto:omar.sleem@kyocera.com)  [amit.kalhan@kyocera.com](mailto:amit.kalhan@kyocera.com) | |

## Outlook of the potential issues (for information only)

|  |  |  |
| --- | --- | --- |
| Issue list | NW-sided model | UE-sided model |
| Configuration for Set A and Set B   * Spatial related information * [Time related information] | Agreement  For network-sided AI/ML model for BM-Case1 and BM-Case2,   * support using existing CSI framework for configuration of Set A as the starting point * support using existing CSI framework for configuration of Set B as the starting point * Note: Purpose, such as above “For NW-sided model, for BM-Case1 and BM-Case2” and “Set A” and “Set B”, will not be specified in RAN 1 specifications | Conclusion  For UE sided model at least for inference, for measurement, the configuration of Set B,   * take the current CSI framework as the starting point   Agreement  For UE-sided model at least for BM Case-1, *CSI-ReportConfig* is used for the configuration of inference results reporting   * FFS on the details in the *CSI-ReportConfig*, at least considering:   + Alt 1: one *CSI-ResourceConfigId* is configured for Set B     - FFS: how UE can determine the information about set A   + Alt 2: one *CSI-ResourceConfigId* is configured for both Set A and Set B     - FFS: How to configure resource set(s) for Set A and Set B in *CSI-ResourceConfig*   + Alt 3: two *CSI-ResourceConfigId* s are configured for Set A and Set B separately   + Alt 4: one *CSI-ResourceConfigId* is configured for Set B, Set A is configured using separate resource set(s) other than that represented by *CSI-ResourceConfigId*     - FFS: how to configure/indicate separate resource set(s) for Set A   + Note: separate *CSI-ReportConfig* for Set A and Set B are not precluded.   + Note: Not perform measurement for Set A and only perform measurement for Set B subject to the *CSI-ReportConfig*   + FFS on the association between Set A and Set B with or without additional IE   + Other necessary configuration are not precluded. |
| Report for inference | Agreement  For NW-sided model, for inference, in a beam report initiated by network, based on one measurement resource set, support the report of more than 4 beam related information in L1 signaling   * Note: Purpose, such as above “For NW-sided model, for inference”, will not be specified in RAN 1 specifications * FFS on the report content for beam related information * FFS on max number of reported beam related information in one report   Agreement  For NW-sided model, for inference report, at least for BM-Case 1, the content in a beam report in L1 signaling, support   * L1-RSRPs and corresponding beam information of Top M beam(s) with largest M measured value(s) of L1-RSRP(s) of a measurement resource set, where M is configured by gNB * If M = the size of the measurement resource set, the content is all L1-RSRPs and one beam index (i.e., CRI/SSBRI) for the largest measured value of L1-RSRP of a measurement resource set * FFS: L1-RSRPs and corresponding beam information of up to M beams within X dB gap to the largest measured value of L1-RSRP, X and M are configured by gNB, and whether/how to report number of reported beams * FFS on the maximum value of M (where M can be larger than 4) based on UE capability (M may or may not be different for different reporting contents) * FFS on beam information * Note: Purpose, such as above “For NW-sided model, for inference report, at least for BM-Case 1”, will not be specified in RAN 1 specifications | Agreement  For UE-sided model, at least for BM-Case1, for content in the report of inference results, support   * Opt 1: Beam information on predicted Top K beam(s) among a set of beams * Opt 2: Beam information on predicted Top K beam(s) among a set of beams and RSRP of predicted Top K beam(s) among a set of beams * At least K=1 and more, FFS on max value * FFS on beam information * FFS on the definition of predicted Top K beam(s) * FFS on definition of reported RSRP when applicable * FFS on other information in the report with potential down selection among the following options * Opt 3: Beam information on predicted Top K beam(s) among a set of beams and probability information of predicted Top K beam(s) among a set of beams   + FFS on the quantization method of probability information   + Probability information is the probability of the beam to be the Top 1 or Top K beam * Opt 4: Beam information on predicted Top K beam(s) among a set of beams, RSRP of predicted Top K beam(s) among a set of beams, and confidence information of the RSRP   + FFS on definition of reported RSRP   + FFS on the definition and quantization method of confidence information * Other options are not precluded.   where the set of beams is Set A, i.e., the beams for UE prediction.  Agreement  For report content of inference results for UE-sided model for BM-Case 1, for the RSRP ofpredicted Top K beam(s) in the report of inference results, when applicable, further study the following options:   * Option A: Predicted RSRP * Option B: Predicted RSRP, if the beam is not configured for corresponding measurement, and measured L1-RSRP if the beam is configured for corresponding measurement * Where the predicted RSRP is based on AI/ML output * Note: Support both Option A and Option B is not precluded.   Agreement  For UE-side AI/ML model inference, for BM-Case2, support to report inference results of N(N>=1, FFS on N) future time instance(s) in one report   * wherein information of inference results of one time instance is as in one report for BM-Case 1   + Note: overhead reduction is not precluded * FFS on details   Agreement  For report content of inference results for UE-sided model for BM-Case 2, the RSRP ofpredicted beam(s) in the report of inference results, is the predicted RSRP, where the predicted RSRP is based on AI/ML output |
| Report for training | FFS | NA |
| L1 signalling general | Agreement  At least for NW sided model, for the quantization of a reported L1-RSRP value at least for the report in L1 signaling, support   * Support differential L1-RSRP reporting with legacy quantization step and range   + FFS: larger quantization step(s) than the already supported legacy quantization step for differential L1-RSRP and/or for absolute L1-RSRP   + FFS: Smaller range(s) for differential L1-RSRP than the already supported legacy range |  |
| Beam indication | Agreement   * For NW-sided model and for UE-sided model, beam indication is based on unified TCI state framework * FFS on whether/how potential enhancement is needed   NA | |
| Consistency and additional condition | FFS   * Rx assumption | Agreement  Further study, for the consistency of NW-side additional condition across training and inference for UE-sided model for BM-Case 1 and BM Case 2, where the NW-side additional condition may at least impact UE assumption on beams of Set A/Set B:   * Opt1: Based on associated ID (Referring to AI 9.1.3.3)   + FFS on what can be assumed by UE with the same associated ID across training and inference   + FFS on how associated ID is introduced, e.g., within CSI framework, or outside of CSI framework * Opt 2: Performance monitoring based   + FFS details * Other options are not precluded. |
| Performance monitoring  (Including report for inference) | Metrics and Procedures | Agreement  For BM-Case1 and BM-Case2 with a UE-side AI/ML model:   * Support Type 1 performance monitoring, including the following two options:   + Option 1 (NW-side performance monitoring):     - UE sends a report to NW (for the calculation of performance metric at NW)       * Measurement results from resource set for monitoring, e.g., L1-RSRP and/or RS index is supported as the content of the report       * FFS on other contents     - The report is at least configured/triggered by NW     - Note: this may or may not have additional spec impact   + Option 2 (UE-assisted performance monitoring):     - UE calculates performance metric(s)       * FFS how to report and what to report   + FFS whether to trigger the report based on event(s) for Option 1 and/or Option 2 * FFS Type 2 performance monitoring |

# Consistency for additional condition for UE-sided model

|  |
| --- |
| Agreement  Further study, for the consistency of NW-side additional condition across training and inference for UE-sided model for BM-Case 1 and BM Case 2, where the NW-side additional condition may at least impact UE assumption on beams of Set A/Set B:   * Opt1: Based on associated ID (Referring to AI 9.1.3.3)   + FFS on what can be assumed by UE with the same associated ID across training and inference   + FFS on how associated ID is introduced, e.g., within CSI framework, or outside of CSI framework * Opt 2: Performance monitoring based   + FFS details * Other options are not precluded. |
| RAN 1 #117  Working Assumption for SI part  Regarding the associated ID for Rel-19, the UE assumes that NW-side additional conditions with the same associated ID are consistent at least within a cell   * FFS: whether/how UE assumption can be applicable for multiple cells (including the feasibility study) |

### Summary from the contributions

|  |  |
| --- | --- |
| Company | View |
| CMCC[5] | Proposal 26: Regarding to the application range of associated ID,   * different gNB vendors use different associated ID set. * for the same gNB vendor, associated ID is global ID or is configured per area.   Proposal 27: Regarding to the UE assumptions behind the same associated ID,   * different gNB vendors do not use the same associated ID. * for the same gNB vendor, UE assumptions including the same transmission filter, the same order of resources (corresponding to beams) for Set A/B, the same Tx power of gNB, the same antenna height and down tilt, the same deployment scenarios.   Proposal 28: For the consistency of NW-side additional condition across training and inference for UE-sided model for BM-Case 1 and BM Case 2, option 1 and option 2 are supported:   * Opt1: Based on associated ID (Referring to AI 9.1.3.3)   + associated ID is introduced within CSI framework and configured per CSI-ReportConfig. * Opt 2: Performance monitoring based |
| Intel [6] | Proposal 28: For a UE-sided AI/ML model, for consistency between training and inference, an ID can be assigned to set A/B configuration which can be assumed to indicate that QCL (at least QCL-D) assumptions for the beams configured in the set and their ordering within the set are consistent when the same ID is used.  Proposal 29: For a UE-sided AI/ML model, for consistency between training and inference, the scope of an Associated ID is limited to within a serving cell.  Proposal 30: For a UE-sided AI/ML model, for consistency between training and inference, performance monitoring-based approaches should be deprioritized. |
| ZTE [7] | Proposal 35: Similar properties of a DL Tx beam set mean that similar beam properties are used for the transmission of the RS resource set and the physical beam corresponding to any entry of the RS resource set remains similar across training and inference, irrespective the resource ID to be used.  Proposal 36: For associated ID based consistency approaches, first clarify the scope of the associated ID before touching upon the configuration details. |
| Ericsson[8] | 1. Proposal 1: For UE-sided models, associated ID comprises a NW TX beam ID, and is configured in the same manner as the TCI states, that is:    * Periodic measurements: Beam IDs are indicated part of NZP-CSI-Resource    * aPeriodic measurements: Beam IDs are part of CSI-AssociatedReportConfigInfo    * Semi-persistent measurements: Beams ID are part of the MAC CE CSI-RS activation message 2. Proposal 2: For UE-sided models, associated ID can be used by UE for categorizing the measurements (as model inputs/labels during training), and associated ID can also be used in the inference operation.    * For example, UE can assume the similar properties of a DL Tx beam with the same associated ID    * FFS: Naming of beam-management specific associated ID (e.g. associated ID is labelled as “beam ID”) 3. Proposal 3: For UE-sided models, associated ID comprises a NW TX beam ID and is cell specific as a starting point    * FFS: Need for beam ID to be valid over multiple cells 4. Proposal 17 For UE-sided models, for the consistency of NW-side additional condition across training and inference addressed via performance monitoring, consider real-time monitoring and study the feasibility with the following aspects as a starting point:    * Frequency of monitoring procedure    * Overhead for monitoring procedure    * Accuracy of monitoring procedure    * Details of monitoring procedure |
| Vivo [9] | 1. Proposal 1: For inference, for UE-side model, it can be assumed that same transmission filter for same resource IDs or through other rules (e.g. same ordering in a list) in Set A/Set B is consistent from training to inference if UE receives the same associated ID. 2. Proposal 2: For inference, for UE-side model, support that associated ID representing NW-side additional conditions is provided to UE to ensure consistency between training and inference, as well as to address NW-side proprietary information disclosing issue. 3. Proposal 3: For inference, for UE-side model, along the line of the agreed working assumption on the consistency issue in AI 9.1.3.3, support to prioritize discussion on how UE-side AI model can work with local associated ID and study how to reduce UE side complexity on managing local associated IDs. 4. Proposal 4: For inference, for UE-side model, suggest to further study on multi-cell AI model with local associated ID, including both the case that the model is developed for multi-cell and the case when UE moves to a cell that UE has not collected data before. 5. Proposal 5: For inference, for UE-side model, further study how to determine consistency for a UE sided AI model for multi-cell AI model, e.g. local associated ID can be mapped to an assistant information (cell ID) during data collection. |
| OPPO [10] | Proposal 21: To ensure consistency between Set A and Set B across training and inference, support (Opt1) the associated ID (from A.I. 9.1.3.3) as its key solution.  Proposal 22: With the associated ID, study what UE can assume from the aspects of Tx beams in Set B and/or Set A across training and inference phases.  Proposal 23: For UE-side model training, support to configure Set B (as model inputs) and Set A (as model labels) using existing CSI framework.  Proposal 24: For UE-side model training, support to configure an associated ID along with the configuration of Set B and Set A. |
| Fujitsu [11] | Proposal 32:   * Regarding the consistency across training and inference, the option based on performance monitoring is preferred. |
| CATT [12] | Proposal 4: For Opt 1 of the consistency of NW-side additional condition across training and inference, the applicable range and the feasibility of the associated ID can be at least within a cell.  Proposal 5: For Opt 2 of the consistency of NW-side additional condition across training and inference, a common dataset for performance monitoring can be used.  Proposal 6: For Opt 2 of the consistency of NW-side additional condition across training and inference, gNB can provide the pre-condition information with preserving privacy information to UE, e.g. the number of vertical beams and the number of horizontal beams of the Tx beam codebook, for avoiding unnecessary performance monitoring. |
| Lenovo [14] | Proposal 4: Support the consistency between training and inference for UE side model by combining an associated ID and performance monitoring. |
| InterDigital[16] | Proposal 18: Support both Opt.1 (based on an associated ID) and Opt.2 (performance monitoring based) for the consistency of NW-side additional conditions. |
| Panasonic [17] | Proposal 7: Support mapping/association of beams within Set A and beams within Set B based on QCL relationship.  Proposal 8: Support to apply concept of “associated IDs” for multiple cells for ensuring consistency of NW-side additional condition across training and inference for UE-sided model for BM-Case 1 and BM Case 2.  Proposal 9: Support to determine “associated IDs” within a NW operator (or a PLMN) to preserve proprietary information. |
| Nokia [19] | Proposal 13: For beam prediction use-cases, RAN1 shall support a solution to ensure consistency between training and inference regarding NW-side additional conditions.  Proposal 14: For beam prediction use cases, to ensure consistency between training and inference regarding NW-side additional conditions, consider introducing an identifier (associated ID) in the CSI-RS resource configuration or measurement resource sets defined within the CSI-RS resource configuration.   * Assigning identifiers (associated ID) shall be left to NW-implementations.   Proposal 15: For beam prediction use cases, to ensure consistency between training and inference regarding NW-side additional conditions, RAN1 spec may define associated ID as the following,   * Associated ID(s) indicated for the data collection may be used by the UE to categorize data samples, where the UE can assume similar properties (aspects other than the configuration parameters) of a DL Tx RS resource set (CSI-RS/SSB resource set) with the same associated ID. The associated ID can also used in the inference operation for alignment (to indicate properties of a DL Tx RS resource set used in the inference). * FFS: whether spec needs to define the details related to similar properties.   Proposal 16: For beam prediction use cases, the performance monitoring/assessment framework shall ensure consistency between training and inference regarding NW-side additional conditions, further discuss the following options,   * Option 1: UE-sided model assessment in a NW-transparent manner (e.g., UE is doing performance assessment to select suitable UE models when supporting beam prediction under different NW assumptions). No spec impacts. * Option 2: UE-sided functionality assessment and reporting the functionality assessment (e.g., as applicable functionality reporting)   + Consider enhancements to enable monitoring of multiple beam prediction related CSI reporting configurations and reporting of applicable CSI report configuration IDs. * Option 3: NW-sided functionality assessment (e.g., NW implementation option where NW selects suitable functionalities based on its own assessments). This option can either be UE-transparent (with no spec impact) or UE-assisted (with some spec impact on RS measurements).   + For UE-assisted operations, consider the changes required on RS measurement and reporting framework. * Option 4: Joint model and functionality assessment by UE and NW. This can be considered as a combination of options 1-3.   + FFS: further discuss details of signalling support. |
| Samsung [21] | Proposal 8. For UE-sided model at least for BM Case-1, for the association between Set A and Set B, introduce DL Tx IDs for the identification of downlink spatial domain transmission filter.   * Each beam in Set A is associated with an DL Tx ID * Each beam in Set B is associated with an DL Tx ID * Note: UE assumes the beams corresponding to the same DL Tx ID shares the same downlink spatial domain transmission filter. * FFS: the relationship between DL Tx ID and associated ID.   Proposal 16. For the consistency of NW-side additional condition across training and inference for UE-sided model for BM-Case 1 and BM Case 2, support the following:   * UE to report the information on the supported/preferred associated ID. * FFS: Other information along with the report of the association ID.   Proposal 17. For the consistency of NW-side additional condition across training and inference for UE-sided model for BM-Case 1 and BM Case 2, support the following:   * NW configuration of associated ID in a CSI-ReportConfig   Proposal 18. For the consistency of NW-side additional condition across training and inference for UE-sided model for BM-Case 1 and BM Case 2,   * At least in a cell level, the UE assumes consistent physical beams characteristics on DL Tx beams of Set A and Set B as well as a consistent mapping, e.g., consistent order of measurement resources |
| Transsion [22] | Proposal 10: Regarding the consistency of NW-side additional condition across training and inference for UE-side model, support Opt1:   * Opt1: Based on associated ID (Referring to AI 9.1.3.3)   + FFS on what can be assumed by UE with the same associated ID across training and inference   + FFS on how associated ID is introduced, e.g., within CSI framework, or outside of CSI framework |
| ETRI [23] | Proposal 1: For the UE-sided model, the UE assumes that the same Associated ID represents a beam with the same direction.  Proposal 2: For the UE-side model, the Associated ID can be configured in the following alternatives:   * CSI-ReportConfig * CSI-ResourceConfig * CSI-ResourceSet   Proposal 3: Support categorizing Associated IDs into cell-specific areas and areas that can be used across multiple cells. |
| CAICT [24] | Proposal 1: Associated ID should be designed based on same Set A/B configuration and per cell group could be considered as starting point. |
| DOCOMO [25] | Proposal 3: Support configuring associated ID within resource related configuration for Set A/B, where the consistency of corresponding Set A or Set B can be assumed.  Proposal 4: Supporting only performance monitoring based approach for consistency across training and inference should be avoided due to UE burden brought by performance monitoring.  Proposal 5: In performance monitoring based approach for consistency over training and inference for Set A/B, the following procedures can be considered.   * Step1: UE reports general beam prediction capability. * Step2: UE receives the message including configuration of Set A/B and request to report beam prediction capability/applicability of corresponding to Set A/B. * Step3: UE check the capability/applicability of corresponding Set A/B via associated ID and/or performance monitoring. * Step4: UE reports the beam prediction capability/applicability of corresponding to Set A and Set B. * Note: the detail/signaling of Step2/4 can be discussed in RAN2. |
| Qualcomm [27] | Proposal 1 For beam prediction for UE-side AI/ML models, consider the following aspects to ensure consistency between training and inference regarding NW-side additional conditions (with regards to Set A, Set B consistency) for inference at UE   * 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. * Beam shape consistency: For each Set A resource, the difference between pointing direction and beamwidth of the physical beam associated with that Set A resource during training compared to pointing direction and beamwidth of the physical beam associated with that same Set A resource during inference should be under predefined tolerances. Similarly, for each Set B resource, the difference between pointing direction and beamwidth of the physical beam associated with that Set B resource during training compared to pointing direction and beamwidth of the physical beam associated with that same Set B resource during inference should be under predefined tolerances.  Proposal 2 For UE-side beam prediction, for the consistency of NW-side additional conditions across training and inference, with regards to FFS on what can be assumed by UE with the same associated ID across training and inference:   * Based on spatial Tx filter: For the same associated ID across training and inference, for each Set A resource, UE can assume that the same spatial TX filter has been utilized by gNB, across training and inference. Similarly, for each Set B resource, UE can assume that the same spatial TX filter has been utilized by gNB, across training and inference.   + Note: a certain tolerance level can be considered for the spatial TX filter used in inference versus training. * Based on QCL relationship: For the same associated ID across training and inference, for each Set A resource, UE can assume that the gNB Tx beams used across training and inference have a certain QCL relationship. Similarly, for each Set B resource, UE can assume that the gNB Tx beams used across training and inference have a certain QCL relationship.   + FFS: definition of such QCL relationships.  Proposal 3 For UE-side beam prediction, for the consistency of NW-side additional condition across training and inference, where the NW-side additional condition may at least impact UE assumption on beams of Set A/Set B, support Opt1: Based on associated ID. Proposal 4 For UE-side beam prediction, for the consistency of NW-side additional condition across training and inference, with regards to FFS on “how associated ID is introduced”, support at least signaling of associated ID within CSI report settings (i.e., CSI-ReportConfig). Proposal 5 For UE-side beam prediction, for the consistency of NW-side additional condition across training and inference, study mechanisms to ensure consistency across different cells.   * Note: As a starting point, study mechanisms to ensure consistency on a per-cell level. |
| Indian Institute of Tech (M), IIT Kanpur [28] | Proposal 1: With associated ID, UE can assume NW-side additional conditions such as beam tilt angle, beam angles (azimuthal and elevation), total number of beams (Set A of beams and/or Set B of beams) of the cell to be consistent during training and inference.  Proposal 2: The associated ID, for AI/ML beam management, could be configured within the CSI framework for both training and inference. |
| Fraunhofer HHI, Fraunhofer IIS [29] | Proposal 1: Functionality LCM associated with a configuration including associated identifiers should be supported for UE-sided models, at least for BM-Case 1.  Proposal 2: Configure a consistency identifier in the Report Config or Resource Config.  Proposal 3: The UE capabilities report may indicate its consistency assumptions for available models or functionalities.  Proposal 4: For UE-side models, support signaling of assistance information from the NW to the UE, at least for BM-Case 1. |
| NTU [30] | Proposal 1: For BM-Case1 and 2, RAN1 to study and consider the model ID free approach which introduces a base station AI/ML model to correct the UE model prediction based on a standardized codebook, e.g., RAN4 testing codebook, and therefore eliminate the model ID indication from the network. The model ID free approach requires the reporting interface with both best beams and measured beams to enable the BS side model to correct codebook mismatch, i.e., the network is allowed to request UE reporting both the best beam based on its prediction and (a subset of) measured beams. |
| ITL [31] | Proposal 7: It can be firstly considered for the association of Set A/B beams to use the current CSI framework as the baseline, including CSI resource, resourceSet, reportConfig, and/or resourceConfig. |
| Huawei/HiSi [33] | Observation 2: For the data collection for the UE-side model, the massive number of impacting factors for consistency between training and inference across cells are difficult to be categorized to associated IDs by NW.  Proposal 11: For the applicable area of the associated ID for the UE-side model, study associated ID subject to cell specific manner as a starting point.  Proposal 12: Associated ID can be used for UE categorizing the measurements, where similar properties of a DL Tx beam set can be assumed with the same associated ID.   * No need to define the physical implication (e.g., beam angle, codebook, beam order, etc.) of the associated ID.   Proposal 13: Consider to indicate the associated ID by reusing the CSI framework, e.g., under CSI-reportConfig, CSI-resourceConfig, etc.  Proposal 14: For the consistency of NW-side additional condition across training and inference for UE-side model, UE side performance monitoring can also be considered. |
| Xiaomi[34] | Proposal 2-8: Support to indicate associated ID to ensure consistency of NW-side additional condition for UE side model.  Proposal 2-9: Support to introduce associated ID within CSI framework per CSI-reportconfig or per resource set.  Proposal 2-10: With same associated ID, UE can assume the consistency of the following NW side additional conditions between training and inference for UE side model.   * + Set B/ set A configuration     - The number of beams in set B     - The number of beams in set A     - The order of beams in set B     - The order of beams in set A     - Pattern of set B       * Contiguous beams or non-contiguous beams in set B       * Fixed or random or preconfigured patterns   + Time window configuration (BM case 2 only)     - The number of history measurement time instance     - The number of predicted future time instance   + Deployment     - Transmission power and antenna height   Proposal 2-11: Support cell group specific associated ID. |
| Kyocera [35] | Proposal 10: For ensuring consistency in network-side additional conditions across training and inference phases for UE-side models in BM Case 1 and BM Case 2, where network-side conditions may impact UE assumptions on beams of Set A/Set B, the following assumptions should be considered by a UE with the same associated ID during both training and inference stages:   * Coverage assumptions:   + - Beam properties: Beam shape, width, and side lobe levels should remain consistent.     - Set Sizes: The sizes of Sets A and B should be similar. * Resource assumptions: The arrangement/indexing of resource indices should be consistent during training and inference stages.   Proposal 11: To ensure consistency of NW-side additional conditions across training and inference for UE-sided models in BM-Case 1 and BM-Case 2, where these conditions may impact UE assumptions on beams of Set A/Set B, we propose defining the associated ID(s) within the current CSI framework. Specifically, introduce the associated ID(s) as a new IE within the RRC messaging of nzp-CSI-RS-ResourceSet and csi-SSB-ResourceSet.  Proposal 12: For A UE side AI/ML model, in BM Case 1 and BM Case 2, to ensure consistency of the NW-side additional conditions across training and inference using an associated ID, RAN1 should discuss how the associated ID can be defined globally among different gNBs within the network. |
| LGE [36] | Observation #3: In current specification, UE cannot assume that Tx beam applied for a NZP CSI-RS resource will be unchanged in different transmission instances.  Observation #4: In current specification, different resource ID does not always mean different Tx beam.  Proposal #18: To address the consistency issue for BM use cases, introduce the following two types of indicators, e.g., so-called associated ID:   * An indicator to ensure same Tx filter in different Tx time instances within a CSI-RS resource * An indicator to inform same Tx filter in different CSI-RS resources |
| NVIDIA [37] | Proposal 9: For AI/ML based beam prediction in spatial/time domain, introduce specification support for additional conditions to include them into model description information during model identification. |
| NEC [38] | Observation 1: The timing information of Set B measurements and Set A predictions is crucial for UE to determine the model to be used for inference.  Proposal 2: In addition to information of beams in Set A and Set B, the timing information of Set B measurements and Set A predictions should be specified as conditions for BM-Case1 and BM-Case2.  Proposal 3: Support associated ID for the consistency of NW-side additional conditions across training and inference for UE-sided model for BM-Case 1 and BM Case 2, where the NW-side additional conditions may at least impact UE assumption on beams of Set A/Set B.  − Associated ID shall at least indicate the site/cell specific variables used for model training like antenna configuration.  − NW-side additional conditions with the same associated ID can be consistent across multiple cells. FFS for which cells NW-side additional conditions can be assumed consistent.  Proposal 4: Support a reference beam pattern to ensure the consistency.  Proposal 5: Support a calibration procedure to ensure the consistency.  Proposal 6: Support to configure the associated ID(s) also for AI/ML performance monitoring. |
| MTK [39] | Proposal 1: For the consistency of NW-side additional condition across training and inference, for Opt1, UE assumes NW maintains the necessarily “beam related properties” across training and inference that will impact model inference performance. The details of “beam related properties” is determined by NW.  Proposal 2: For configuring the associated ID, RAN1 consider the following two directions:   * Alt1: associated ID is configured within the CSI-framework   + AltA: associated ID is configured in CSI report configuration   + AltB: associated ID is configured when NW triggers/activates the CSI reports * Alt2: associated ID is configured outside of the CSI-framework, where the associated ID is assumed to be shared across all the csi-report and resource configurations   + FFS: details (e.g., NW indicates the associated ID by a new DCI format, then UE switches to the new model after receiving the corresponding DCI) * Note: whether/how to prevent multiple associated IDs are configured at the same time   Proposal 3: To save the number of configuration IDs, for configuring the associated ID within the CSI framework, RAN1 does not consider configuring the associated ID within the CSI Resource/ResourceSet/ ResourceConfig configurations.  Proposal 4: Support the specification of a model switch time, which is defined as the time required for UE to switch to a new model after receiving a new associated ID  Proposal 5: For a global associated ID, RAN1 study the following:   * how to limit the number of possible associated IDs * how to reduce signaling overhead of associated ID   Proposal 6: For a per cell associated ID, RAN1 study the following:   * how to use these per cell associated ID to indicate UE to update/retrain a model * how to use these per cell associated ID to indicate UE to fine-tuning a model   Proposal 33: For BM-Case1 and BM-Case2 with a UE-side AI/ML model, for the BM-specific conditions regarding “information regarding model inference”, consider at least 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)   Proposal 34: For BM-Case1 and BM-Case2 with a UE-side AI/ML model, for the BM-specific conditions regarding “performance monitoring”, consider at least for the following sub-conditions,   * conditions on performance metrics * conditions on the detectable events |
| APPLE [40] | Proposal 6-1: The associated ID in assisted information needs to be PLMN unique, and core network or O&M is involved in assigning/managing the associated ID.  Proposal 6-2: The associated ID/assistance information/ (e.g., dataset ID/model ID), if assigned by higher layer is embedded as part of reference signal configuration.  Proposal 6-3: the introduction of “associated ID” should provide a means for UEs to collect training data and for the UE side to train AI/ML models with reasonable effort. Beyond PLMN and PLMN-specific-ID, consider a Tag field to further reduce data collection/data collection effort.  Proposal 6-4: For CSI report of UE side model inference output, SSBRI/CRI based on set A configuration for training or performance monitoring is reported. The set A RS configuration with an associated ID during inference report and performance monitoring, should have QCL-type D relationship with the set A RS configuration with the same associated ID during training. |
| KT [41] | Proposal 1. RAN1 confirm that the associated ID can be considered as a solution to ensure consistency of NW-side additional condition between training and inference for UE-sided models with functionality identification.  Proposal 2. The associated ID(s) be introduced within the CSI framework at least for model training.  Proposal 3. Support the performance monitoring of the inactive model(s) before activation of the applicable model(s)/functionality(es). |
| Meta [42] | **Proposal 2: Associated ID should be used for consistency of NW-side additional conditions for a UE side model.** |

### Issue #1.1: Associated ID for UE sided model

##### (FL0) Proposal 1.1:

For UE sided model in beam management, introduce associated ID to ensure consistency across training and inference, where the associated ID can be configured for UE categorizing the data samples for training (as model inputs/labelling), and the associated ID can also be configured in the inference operation for ensuring consistency

* UE can assume the *similar properties* of a DL Tx beam set/list associated with the same associated ID
  + FFS: whether/how to define *similar properties* of the DL Tx beam set/list, e.g.
    - Similar physical beam characteristics of each beam [with the same index ordered in Set A or Set B]
    - consistent mapping [of measurement resources] in Set A of beams, and Set B of beams
  + Note: There is no restriction on the physical implication of the associated ID.
* The associated ID at least can be configured within CSI framework (with RS resource configuration)
  + FFS on details
  + FFS on whether to configure/indicate the associated ID via other signal(s) and/or in other procedure(s)/framework(s)
* The associated ID can be used to ensure consistency at least on a per-cell level.
  + FFS on whether/how to ensure consistency across different cells.

|  |  |
| --- | --- |
| Supported companies | FL |
| Companies | View |
| FL | Regarding on the “similar properties” and the order. In my opinion, what we need is the same physical beam characteristics of the beams with same index ordered in Set A (#n beam in Set A). other than the configured resources, or some ID for the resources (e.g., NZP-CSI-RS-ResourceId) for #n beam in Set A. therefore, the second sub-sub-bullet highlighted in yellow can be deleted. |
| NTT DOCOMO | We suggest deleting the second sub-sub-bullet highlighted in yellow, as the first sub-sub-bullet is sufficient. |
| OPPO | We are in general supportive to this FL proposal.  Regarding the highlighted as yellow part, we think it’s no harm to keep the consistent mapping of Set A and Set B to make sure the ordering of inputs of model. And the measurement resource in the bracket can be removed, since what can be configured as Set A is not determined yet. If that’s the case, could we suggest the following wording for consideration?   * + FFS: whether/how to define *similar properties* of the DL Tx beam set/list, e.g.     - Similar physical beam characteristics of each beam [with the same index ordered in Set A or Set B]     - consistent mapping ~~[of measurement resources]~~ in Set A of beams, and Set B of beams |
| HW/HiSi | 1) The linkage between UE and gNB on the interpretation of associated ID is a weak linkage: gNB may or may not adjust the mapping order/beam shape/codebook of Set A/Set B between training and inference, and as long as it feels such change is too minor to impact UE perception, it can assign with the same associated ID.  2) The mapping order does not need to be specified for the associated ID – it can be specified for how to use Set A and/or Set B for the beam measurement part, such as “UE assumes beams in Set A/Set B are mapped from LSB to MSB of the resource set”  Therefore, the FFS on the interpretation of the properties is removed.  For UE sided model in beam management, introduce associated ID to ensure consistency across training and inference, where the associated ID can be configured for UE categorizing the data samples for training (as model inputs/labelling), and the associated ID can also be configured in the inference operation for ensuring consistency   * UE can assume the *similar properties* of a DL Tx beam set/list associated with the same associated ID   + FFS: whether/how to define *similar properties* of the DL Tx beam set/list, e.g.     - Similar physical beam characteristics of each beam [with the same index ordered in Set A or Set B]     - consistent mapping [of measurement resources] in Set A of beams, and Set B of beams   + Note: There is no definition ~~restriction~~ on the physical implication of the associated ID. * The associated ID at least can be configured within CSI framework (with RS resource configuration)   + FFS on details   + FFS on whether to configure/indicate the associated ID via other signal(s) and/or in other procedure(s)/framework(s) * The associated ID can be used to ensure consistency at least on a per-cell level.   + FFS on whether/how to ensure consistency across different cells. |
| Fujitsu | If the associated ID is for consistency at cell level, it means there would be a lot of number of conditions for the UE to check considering the number of cells.  It’s still not clear to us how to make it work at per-cell level. |
| New H3C | OK in general |
| TCL | Agree. The associated ID should reflect the alignment of NW additional conditions in training and inference. |
| ETRI | We are generally fine with this proposal. |
| Xiaomi | Support the main bullet of the proposal.  As for the ‘similar physical beam characteristics’, it is better to clarify the definition of the ‘similar physical beam characteristics. Does it mean the beam width, the direction, beam shape and Tx beam angle? As for the direction, it is a global direction or a local/relative direction? We think it is better to assume the same local/relative direction, e.g., the reference is the panel.  In addition, regarding to ‘[with the same index ordered in Set A or Set B]’, the index is the ‘beam index’ or ‘resource index’? we think ‘beam index’ is more reasonable, but there is no beam index in Set A or Set B.  As for the yellow part, it can be removed and we can try to make the first sub-bullet clearer.  And we also have some concerns on the complexity of the per cell level. We prefer cell group level. |
| ZTE | For the first bullet, discussions on next step details of similar properties is not necessary. There are many potential factors (e.g., beam shape, antenna down tilt angel, antenna height) may cause model performance degradation, and which factor or to what extent should be considered for associated ID is strongly related to the model's generalization capability. Therefore, it would be difficult to give a concrete definition of associated ID that all UEs can assume. Instead, all UEs can assume that the associated ID is indicated for data categorization to ease UE-side data collection for model training/inference specific to AI beam management.  For the second bullet, compared with configuring the associated ID in the RS resource configuration, it would be more beneficial to configure it in the triggering or activation signaling of resource set. Specifically, if the associated ID is configured in the resource set configuration, each time the beneath beam shape changes a new resource set configuration should be used, which causes a lot of configuration redundancy. On the contrary, the same resource set configuration can be maintained if the associated ID in configured in the triggering or activation signaling.  For the third bullet, there would be many feasibility issues related to associated ID to be applied to multiple cells.  Therefore, we suggest the following revisions to the bullets.   * UE can assume the *similar properties* of a DL Tx beam set/list associated with the same associated ID   + FFS: whether/how to define *similar properties* of the DL Tx beam set/list, e.g.     - Similar physical beam characteristics of each beam [with the same index ordered in Set A or Set B]     - consistent mapping [of measurement resources] in Set A of beams, and Set B of beams   + Note: There is no restriction on the physical implication of the associated ID. * The associated ID at least can be configured within CSI framework ~~(with RS resource configuration)~~   + FFS on details   + FFS on whether to configure/indicate the associated ID via other signal(s) and/or in other procedure(s)/framework(s) * The associated ID can be used to ensure consistency at least on a per-cell level.   + FFS on whether/how to ensure consistency across different cells (including the feasibility study) |
| Qualcomm | As long as we ensure that for each resource index in Set B (e.g., resource index #1) the physical beam shape associated with that resource index across training and inference is *similar*, we do not need to discuss consistent indexing/ordering separately, as the consistent indexing/ordering is automatically ensured in this way. Similar argument is true for each resource index in Set A. with that said, the first FFS is updated as follows. What is meant by first note is not clear, so suggest removing it.  For the second bullet, the associated ID can be within CSI framework but configured within CSI *report* configuration. As discussed before, we can have different granularities in terms of options for configuring associated ID, and within the CSI framework, the coarsest granularity would be CSI report configuration. As mentioned by FL, details can be FFS at this point.  Updated Proposal 1.1  For UE sided model in beam management, ~~introduce~~ support associated ID to ensure consistency across training and inference, where the associated ID can be configured for UE categorizing the data samples for training (as model inputs/labelling), and the associated ID can also be configured in the inference operation for ensuring consistency   * UE can assume the *similar properties* of a DL Tx beam set/list associated with the same associated ID   + FFS: whether/how to define *similar properties* of the DL Tx beam set/list, e.g.     - Similar physical beam shape characteristics of each beam ~~[with the same index ordered in Set A or Set B~~ for each resource index in Set A and each resource index in Set B, across training and inference~~]~~, e.g., similar spatial Tx filter, similar quasi-collocation properties, etc.     - consistent mapping [of measurement resources] in Set A of beams, and Set B of beams   + Note: There is no restriction on the physical implication of the associated ID. * The associated ID at least can be configured within CSI framework ~~(with RS resource configuration)~~   + FFS on details   + FFS on whether to configure/indicate the associated ID via other signal(s) and/or in other procedure(s)/framework(s) * The associated ID can be used to ensure consistency at least on a per-cell level.   + FFS on whether/how to ensure consistency across different cells.   Note: ensuring consistency across different cells is crucial for the operation of UE-side models |
| InterDigital | Fine for the main bullet, but have some issues especially for the highlighted parts below.  1. What is the intention of no restriction on the physical implication? Difficult to understand the intention.  2. As we believe that the associated ID should be commonly used for different functionality, we don’t support configuring the associated ID within CSI framework. (FL0) Proposal 1.1: For UE sided model in beam management, introduce associated ID to ensure consistency across training and inference, where the associated ID can be configured for UE categorizing the data samples for training (as model inputs/labelling), and the associated ID can also be configured in the inference operation for ensuring consistency   * UE can assume the *similar properties* of a DL Tx beam set/list associated with the same associated ID   + FFS: whether/how to define *similar properties* of the DL Tx beam set/list, e.g.     - Similar physical beam characteristics of each beam [with the same index ordered in Set A or Set B]     - consistent mapping [of measurement resources] in Set A of beams, and Set B of beams   + Note: There is no restriction on the physical implication of the associated ID. * The associated ID at least can be configured within CSI framework (with RS resource configuration)   + FFS on details   + FFS on whether to configure/indicate the associated ID via other signal(s) and/or in other procedure(s)/framework(s) * The associated ID can be used to ensure consistency at least on a per-cell level.   + FFS on whether/how to ensure consistency across different cells. |
| Ericsson | In general fine with the proposal. We should also cover the scenario when the associated ID is supported on a resource-level (one associated ID per beam), it is not clear if this is the case with current wording. Moreover, *similar properties* are a controversial topic, propose to keep it open. Hence, our view is to change to the following bullet according to below,  ……   * UE can assume the *similar properties* of DL Tx beam(s) ~~set/list~~ ~~associated~~ with the same associated ID   + The DL Tx beam(s) can comprise a single beam, a set of beams, or a list of beams.   + FFS: whether/how to define *similar properties*   ……. |
| LG | Generally fine with the direction. Some comments from our side as below.   * Prefer to discuss what can be assumed by UE with the same associated ID firstly. Then, we can easily decide where to configure the associated ID within CSI framework, e.g., resource set-level, resource level.   + The meaning of“the similar properties of a DL Tx beam” is ambiguous. We prefer to use “the same DL Tx filter” as more specification wording like repetition=ON description.   + Then, we can further discuss configuration granularity of the associated ID. If we have the associated ID per resource level, UE assume the same DL Tx spatial filter of the corresponding resources under same association ID for different resources, and we don’t need to care about the ordering of beam index in Set A or Set B. * The exact name of ‘associated ID’ in specification should be FFS. In our view, the exact name should be decided to represent the exact meaning of the ID well. Different ID name can be used for different use case, e.g. for BM and for Positioning. |
| CATT | In our understanding, associate ID can be used to categorize data sample in training phase, and help UE select the appropriate model/functionality in inference phase across different times within the cell. Also, there is a working assumption in AI 9.1.3.3 that same associated ID can be used at least within a cell, and the case of multiple cells is still FFS. In this case, we suggest to change the main bullet as following:  For UE sided model in beam management, introduce associated ID to ensure consistency across training and inference within a cell, where the associated ID can be configured for UE categorizing the data samples for training (as model inputs/labelling), and the associated ID can also be configuredin the inference operation for ensuring consistency  In additional, It is not clear what kind of difference can be viewed as *similar properties*. In our view, it is unnecessary to discuss the exact NW-side additional condition behind the associated ID, we suggest the following:  UE can assume the functionality can be applied when the associated ID configured in inference phase is the same with that during training phase. |
| vivo | We are generally fine with the direction, and we think this is one of the most important issues to be addressed in this meeting.  On the detailed wording, we think the highlighted part (consistent mapping [of measurement resources] in Set A of beams, and Set B of beams) in the FFS subbullet is one key issue here. The issue is whether gNB will assign different sets of resources (i.e., resource IDs) for training and inference from NW flexibility perspective. If this is flexibility is needed for gNB, the spec shall be clear about the yellow part, i.e., the mapping or ordering of resources in the resource sets of training and inference. On the other hand, if this flexibility is not allowed, the spec should clarify that same sets of resource IDs should be maintained for training and inference in order not to confuse UE’s implementation. |
| Fraunhofer | We support the proposal in general. With regards to the bullet point on consistent mapping, we think that this property can be derived from the previous bullet. However, there is no harm to keep it as clarification. |
| Nokia | General direction is ok. Some edits on the proposal is as below.  Updated proposal (comments in blue)  For UE sided model in beam management, introduce associated ID to ensure consistency across training and inference, where the associated ID can be configured for UE categorizing the data samples for training ~~(as model inputs/labelling)~~, and the associated ID can also be configured in the inference operation for ensuring consistency %Nokia : listing model inputs/labelling does not seem to be critical   * UE can assume the *similar properties* of a DL Tx beam set/list associated with the same associated ID   + FFS: whether/how to define *similar properties* of the DL Tx beam set/list~~, e.g.~~     - Similar physical beam characteristics of each beam [with the same index ordered in Set A or Set B] %Nokia : e.g., does not seem to be critical. We can discuss those when “whether” is concluded.     - consistent mapping [of measurement resources] in Set A of beams, and Set B of beams %Nokia : Let’s keep the proposal simple first.   + Note: There is no restriction on the physical implication of the associated ID. * The associated ID at least can be configured within CSI framework (with RS resource configuration)   + FFS on details   + FFS on whether to configure/indicate the associated ID via other signal(s) and/or in other procedure(s)/framework(s) %Nokia : Let’s first discuss “where to place this associated ID in CSI framework”. * The associated ID can be used to ensure consistency at least on a per-cell level.   + FFS on whether/how to ensure consistency across different cells. |
| Panasonic | Regarding “similar properties”, we think it is not necessary to introduce this new terminology because we already have a terminology of “NW-additional conditions”. Therefore, we propose to revise this bullet as   * When reception of the same associated ID, UE can assume the same configuration of NW-side additional conditions that used for a DL Tx beam set/list.   Regarding 2nd FFS point of the 2nd sub-bullet, it is not necessary because the 1st sub-bullet can cover it.  In general, we propose the following (update is highlighted as red color):  **Proposal 1.1**: For UE sided model in beam management, introduce associated ID to ensure consistency across training and inference, where the associated ID can be configured for UE categorizing the data samples for training (as model inputs/labelling), and the associated ID can also be configured in the inference operation for ensuring consistency   * When reception of the same associated ID, UE can assume the same configuration of NW-side additional conditions that used for a DL Tx beam set/list. * UE can assume the *similar properties* of a DL Tx beam set/list associated with the same associated ID   + FFS: whether/how to define *similar properties* of the DL Tx beam set/list, e.g.     - Similar physical beam characteristics of each beam [with the same index ordered in Set A or Set B]     - consistent mapping [of measurement resources] in Set A of beams, and Set B of beams   + Note: There is no restriction on the physical implication of the associated ID. * The associated ID at least can be configured within CSI framework (with RS resource configuration)   + FFS on details   + FFS on whether to configure/indicate the associated ID via other signal(s) and/or in other procedure(s)/framework(s) * The associated ID can be used to ensure consistency at least on a per-cell level.   + FFS on whether/how to ensure consistency across different cells. |
| MediaTek | For the 1st bullet, RS resource configuration is not the only place that the associated ID can be configured at. In our view, associated ID can be configured in CSI-report level or when a report is triggered/activated. We suggest removing “with RS resource configuration”. Also we think it is important to discuss how to limit the number of associated IDs that can be configured simultaneously to UE. If too many associated IDs are configured to UE simultaneously, UE might need to run multiple models to generate the report.  For the 3rd bullet, we think it can be discussed in a separate proposal.  Therefore, we suggest the following wording change:   * The associated ID at least can be configured within CSI framework ~~(with RS resource configuration)~~   + FFS on details   + FFS on whether to configure/indicate the associated ID via other signal(s) and/or in other procedure(s)/framework(s)   + FFS maximum number of associated ID that can be configured simultaneously to UE, subjected to UE capability * The associated ID can be used to ensure consistency at least on a per-cell level.   FFS on whether/how to ensure consistency across different cells. |
| Spreadtrum | We are generally fine with this proposal. For the sub-sub-bullet highlighted in yellow, we think there is confusion for the consistent mapping in Set A of beams, and Set B of beams. We think consistent mapping seems to be the detail for the “Similar physical beam characteristics of each beam”.Thus we support to delete the yellow bullet. |
| CMCC | The association id is to keep the consistency between training and inference. But it is still open whether there is a need to introducing a mapping between Set A and Set B. The Set A beams and Set B beams can share a same set of spatial filter, but it does not mean that we need some kind of spatial filter mapping explicitly between Set A and B.  Besides, mapping of Set A and Set B does not belong to NW side additional condition, thus the highlighted part can be removed. |
| Kyocera | Overall, we are supportive of this proposal and agree with retaining the highlighted section in yellow. This is because the features of the ML model used during training should be applied consistently during inference. In other words, if measurements from resource "i" are used for model input pin "i" during training, then resource "i" must also be used for model input pin "i" during inference. Similar argument can be made for the model outputs. |
| Sharp | Support. |
| Futurewei | We are in principle fine with FL’s proposal. Suggest removing the sub-bullets under “UE can assume the *similar properties* of a DL Tx beam set/list associated with the same associated ID” to make it high level at this stage. |
| NEC | We cannot introduce the word ‘similar’ into spec. Since it is UE assumption, we believe ‘assumed to be the same’ is fine. |
| Lenovo | We are general fine to ensure consistency based on associated ID. Some comments are provided as follows:  Firstly, does the intention of the first bullet is to specify the properties or the additional conditions in the standard?  Secondly, regarding the second bullet. Does the second FFS mean that the associated ID for some CSI framework component can be updated by MAC CE or DCI? Further, the procedure or framework other than CSI framework that need to be indicated/configured with an associated ID is not clear to us. |
| Apple | In general, the FL’s summary is on the right direction. On where the associated ID is configured, we need to see whether for UE-side training data collection whether any CSI report to the NW side is required also. If not, we think the place where the associated ID is configured is not necessarily associated with a CSI report. |

##### (FL1) Proposal 1.1b:

Proposal

* with the same associated ID
  + UE may assume *similar properties* for the same entry in a resource set or a resource set used as Set A across training/inference/[monitoring]
  + UE may assume *similar properties* for the same entry in a resource set or a resource set used as Set B across training/inference/[monitoring]
  + Note: The *properties* at least include downlink spatial domain transmission filter
  + UE may assume the resources within a resource set used as Set A are transmitted with a set of *similar* downlink spatial domain transmission filters
  + Note: There is no restriction on the physical implication of the associated ID.
* Note: Purpose, such as above “across training/inference/[monitoring]”, may or may not be specified in RAN 1 specifications

Proposal A:

* UE may assume the *similar properties* of a DL Tx beam or beam set/list associated with the same associated ID
  + FFS: whether/how to define *similar properties* of a DL Tx beam or beam set/list

Proposal B:

* The associated ID is used to ensure consistency between training and inference regarding NW-side additional conditions for inference at UE
* RAN 1 has no consensus on whether/how to define NW-side additional conditions or UE assumption with the same associated ID

Proposal C:

* with the same associated ID, the measurements based on a resource set (e.g., as Set A or as Set B) can be collected corresponding to the order of entry in the list of the resource set.

|  |  |
| --- | --- |
| Companies | View |
| FL | Let’s try proposal A.  Otherwise, proposal B will be the fall back.  Proposals C is a new wording to ensure the ordering. Which can be combined with propal A or proposal |

TS 38.214:

|  |
| --- |
| The UE can be configured with a list of up to *M TCI-State* configurations within the higher layer parameter *PDSCH-Config* to decode PDSCH according to a detected PDCCH with DCI intended for the UE and the given serving cell, where M depends on the UE capability *maxNumberConfiguredTCIstatesPerCC*. Each *TCI-State* contains parameters for configuring a quasi co-location relationship between one or two downlink reference signals and the DM-RS ports of the PDSCH, the DM-RS port of PDCCH or the CSI-RS port(s) of a CSI-RS resource. The quasi co-location relationship is configured by the higher layer parameter *qcl-Type1* for the first DL RS, and *qcl-Type2* for the second DL RS (if configured). For the case of two DL RSs, the QCL types shall not be the same, regardless of whether the references are to the same DL RS or different DL RSs. The quasi co-location types corresponding to each DL RS are given by the higher layer parameter *qcl-Type* in *QCL-Info* and may take one of the following values:  - 'typeA': {Doppler shift, Doppler spread, average delay, delay spread}  - 'typeB': {Doppler shift, Doppler spread}  - 'typeC': {Doppler shift, average delay}  - 'typeD': {Spatial Rx parameter}  *repetition* in *NZP-CSI-RS-ResourceSet* is associated with a CSI-RS resource set and defines whether UE can assume the CSI-RS resources within the NZP CSI-RS Resource Set are transmitted with the same downlink spatial domain transmission filter or not as described in Clause 5.1.6.1.2. and can be configured only when the higher layer parameter *reportQuantity* associated with all the reporting settings linked with the CSI-RS resource set is set to 'cri-RSRP', 'cri-SINR', 'cri-RSRP- Index', 'cri-SINR- Index' or 'none'. |

### Issue #1.2(on hold): Whether/how to apply performance monitoring-based method to ensure the consistency

TBD

# Performance monitoring (Metrics/events/procedures)

|  |
| --- |
| In TR 38.843  For BM-Case1 and BM-Case2 with a UE-side AI/ML model:  - Type 1 performance monitoring:  - Configuration/Signalling from gNB to UE for measurement and/or reporting  - UE may have different operations  - Option 1 (NW-side performance monitoring): UE sends reporting to NW (e.g., for the calculation of performance metric at NW)  - Option 2 (UE-assisted performance monitoring): UE calculates performance metric(s), either reports it to NW or reports an event to NW based on the performance metric(s)  - Indication from NW for UE to do LCM operations  - Note: At least the performance and reporting overhead of model monitoring mechanism should be considered  - Type 2 performance monitoring:  - Indication/request/report from UE to gNB for performance monitoring  - Note: The indication/request/report may be not needed in some case(s)  - Configuration/Signalling from gNB to UE for performance monitoring measurement and/or reporting  - If it is for UE side model monitoring, UE makes decision(s) of model selection/activation/ deactivation/switching/fallback operation  - Mechanism that facilitates the UE to detect whether the functionality/model is suitable or no longer suitable  Agreement  For BM-Case1 and BM-Case2 with a UE-side AI/ML model:   * Support Type 1 performance monitoring, including the following two options:   + Option 1 (NW-side performance monitoring):     - UE sends a report to NW (for the calculation of performance metric at NW)       * Measurement results from resource set for monitoring, e.g., L1-RSRP and/or RS index is supported as the content of the report       * FFS on other contents     - The report is at least configured/triggered by NW     - Note: this may or may not have additional spec impact   + Option 2 (UE-assisted performance monitoring):     - UE calculates performance metric(s)       * FFS how to report and what to report   + FFS whether to trigger the report based on event(s) for Option 1 and/or Option 2 * FFS Type 2 performance monitoring |

### Summary from the contributions

|  |  |
| --- | --- |
| Companies | Proposals |
| Futurewei [1] | Proposal 1: For Rel-19 AI/ML-based BM, only support UE reporting L1-RSRP and the corresponding CRI/SSBRI for performance monitoring of BM-Case1 and BM-Case2. |
| Tejas[4] | Proposal 6: For BM-Case1 and BM-Case2 with a UE-side AI/ML model, consider Top-K beam prediction accuracy report for UE-assisted performance monitoring.  Proposal 7: For BM-Case1 and BM-Case2 with a UE-side AI/ML model, consider L1-RSRP difference report for UE-assisted performance monitoring.  Proposal 8: For Type 2 performance monitoring for UE-sided model, the UE request for performance monitoring by indicating the AI/ML functionality and the performance metric.  Proposal 9: For Type 2 performance monitoring NW can assign AI/ML functionality and performance metric to the UE |
| CMCC[5] | Proposal 30: Regarding Type 2 monitoring of UE-side AI/ML model, NW may configure a threshold criterion to facilitate UE to perform model monitoring. The specification impact of decision and configuration reporting, and decision acknowledgement mechanism are considered.  Proposal 31: Regarding signaling of data collection for monitoring, L1 signaling is supported.  Proposal 32: Regarding Type 1 option 1 monitoring, of UE-side AI/ML model, KPI is up to gNB. UE report is a single sample-based report. The report content can be measurement results of resource set for monitoring, e.g., L1-RSRP and/or Top 1 RS ID, or Top-K/1 prediction accuracy per sample.  Proposal 33: Regarding Type 1 option 2 monitoring of UE-side AI/ML model, KPI can be single sample-based Top-1/K beam prediction accuracy or multiple sample-based Top-1/K beam prediction accuracy.  Proposal 34: Regarding Type 1 option 2 monitoring of UE-side AI/ML model, NW may configure a threshold criterion or event to facilitate UE to perform model monitoring. The event can be defined as consecutive N1 times of KPI/BLER less than threshold, or accumulated N1 times of KPI/BLER less than threshold within a time period.  Proposal 35: The best beam(s) obtained by measuring beams of a set indicated by gNB is considered as the benchmark/reference for monitoring performance comparison:  o gNB configures one or multiple sets for one or multiple benchmarks/references  Proposal 36: For Type 2 and Type 1 option 2 monitoring of UE-side AI/ML model, monitoring mechanism of multiple benchmarks needs discussion to facilitate UE to perform model monitoring. |
| Intel [6] | Support Alt-1, Alt-2, and Alt-4 for model monitoring metrics:   * Alt.1: Beam prediction accuracy related KPIs, e.g., Top-K/1 beam prediction accuracy. * Alt.2: Link quality related KPIs, e.g., L1-RSRP, L1-SINR, etc. * Alt.4: The L1-RSRP difference evaluated by comparing measured RSRP and predicted RSRP.   For UE-side AI/ML models, for BM-Case 1 and BM-Case 2, for reporting of model monitoring related information for Type 1 monitoring, support:   * Option A: Report the measurement results (e.g. L1-RSRP and/or beam information) of a set of beams   + Baseline assumption: the set of beams is the set A of beams; FFS: subset of set A or smaller set of beams than set A * Option B: Report the beam prediction accuracy related information within a configured time window, e.g.,   + Top-K/1 beam prediction accuracy   + Extended-Top-K/1 beam prediction accuracy such that the “Extended-Top-K/1 beams” include the Top-K/1 measured beams and any additional beams with L1-RSRP within a specified or configured X dB margin, e.g., X = 1 dB * Option C: Report the RSRP difference information between the measured and predicted beams, e.g.,   + For reporting of Top-1 beam, the L1-RSRP difference between the predicted Top-1 beam and the measured Top-1 beam in the set of beams identified for measurements for model monitoring is reported to the network.     - The set of beams can be set A or set B; FFS: other options   + For reporting of Top K beams,     - Alt. C1: the L1-RSRP difference between each of the predicted Top-K beams and the measured Top K beams (respectively) from the set of beams identified for measurements for model monitoring are reported to the network. Alt. C2: the lowest L1-RSRP for the Top K beams (e.g., corresponding to the K-th best beam) for the predicted and measured Top K beams respectively are used to calculate the RSRP difference for reporting as a model monitoring metric.   For network-side AI/ML models, consider UE event triggered model monitoring with periodic or aperiodic measurement on set B of beams.  For UE-side AI/ML models, support UE-event-triggered reporting of model monitoring results for Options 1 or 2.   * Once triggered, UE requests network for resources to report monitoring results. * Triggering events can be defined based on one or more of the following: * (Event-1) The measured Top-1 or Top K beam(s) of Set A and any additional beams with L1-RSRP values within a specified or configured margin X dB compared to the Top K beams, e.g., X = 1 dB and the predicted Top-1 or Top K beam(s) of Set A are different for more than a threshold number of instances within a defined window.   + - FFS: Use of a subset of set A of beams instead of full set A * (Event-2) The measured L1-RSRP of one set of beams is lower than a threshold for more than a threshold number of instances within a defined window.   + - The set of beams includes the Top-1 or Top K predicted beams.   + (Event-4) The L1-RSRP difference between the measured Top-1 or Top K beam(s) of Set A and predicted Top-1 or Top K beam(s) of Set A are larger than a threshold value where the associated beams for RSRP difference are defined using:     - RSRP difference between predicted Top-1 beam and the measured Top-1 beam of a set of beams       * Set of beams is same as set A of beams or FFS: a subset, e.g., set B of beams     - Difference between the lowest L1-RSRP amongst the predicted Top K beams, and the lowest L1-RSRP amongst the Top K beams of a set of beams       * Set of beams is same as set A of beams or FFS: a subset, e.g., set B of beams   For model monitoring for BM-Case-2, further consider if the time window for model monitoring should be identical to observation window for model inferencing.  For UE-side AI/ML model selection/switching, consider network configuration, subject to UE capability, of periodic performance monitoring of non-active models and comparison to the KPI of current models such that an alternate model with potentially better performance can be selected or switched. |
| ZTE [7] | Proposal 27: Support beam prediction accuracy related KPIs (i.e., Alt.1) as the primary performance metric for AI/ML performance monitoring.  Proposal 28: Considering the limited payload size and near-real-time latency requirement, support L1 measurement (CSI reporting) for collecting data to enable Option 1 NW-side performance monitoring.  Proposal 29: For Option 2 (UE-assisted performance monitoring), further consider monitoring report initiated by NW and monitoring report initiated by UE.  Proposal 30: For monitoring report initiated by NW, the report content can be a calculated performance metric or an indication about whether the calculated performance metric is larger than or equal to a configurable threshold.  Proposal 31: Consider UE-initiated monitoring report on the basis of the beam failure recovery mechanisms specified in current specifications.  Proposal 32: Type 2 performance monitoring (i.e., UE-side performance monitoring) can only be supported if the UE is authorized by the NW for functionality or model operations.  Proposal 33: Model/functionality failure detection should be based on monitoring results of several consecutive times within a predefined monitoring window.  Proposal 34: UE reporting based on measurement of Set B can serve as an always-on fallback method to guarantee continuous services quality. |
| Ericsson [8] | Proposal 13 For UE-sided performance metric reporting, at least support beam prediction accuracy (Top-1, Top-3, Top-1/3 within a threshold)  Proposal 14 For UE-sided model performance metric reporting, support statistical prediction performance metrics (e.g., mean, 10th, 50th, 90th percentile) of   * L1-RSRP prediction error, * L1-RSRP error comprising measured difference in the L1-RSRP of the predicted strongest beam, and actual strongest beam * FFS: Number of samples needed for the aggregated report   Proposal 15 For UE-sided model performance metric reporting, consider UE reporting of performance metrics as part of the inference report.  Proposal 16 For NW-side performance metric calculation of UE-sided model (type1, option1), conclude that there is no specification impact specific to enable the performance metric calculation at NW-side |
| Vivo [9] | Proposal 6: For performance monitoring, at least support beam prediction accuracy related KPI and L1-RSRP difference as potential metrics, while probability and confidence information should be deferred after related agreement is achieved in inference phase.  Proposal 7: For performance monitoring, at least support L1 beam reporting for both UE-side model and NW-side model.  Proposal 8: For monitoring report, support report overhead reduction for NW-side performance monitoring of UE-side AI model (i.e. type 1 option 1), as well as for NW-side AI model, e.g., report L1-RSRPs and corresponding beam information of up to M beams within X dB gap to the largest measured value of L1-RSRP.  Proposal 9: For monitoring report, support to report measured L1-RSRP of indicated beam(s), e.g. current beams indicated by DCI, with measured L1-RSRP of top-k beams for both NW-side performance monitoring of UE-side AI model (i.e. type 1 option 1) and NW-side AI model.  Proposal 10: For performance monitoring, for UE-side model, support Option B and Option C, i.e. report the beam prediction accuracy related information and report the RSRP difference information between the measured and predicted, for further study, while Option D and Option E, i.e. probability and confidence information, should be deferred after related agreement is achieved in inference phase.  Proposal 11: For performance monitoring, for UE-side model, support following options for further study on what/how to report,   * Opt1: report metric * Opt2: report event * Opt3: based on event, report event and metric(s) * Other options cannot be precluded   Proposal 12: For performance monitoring, for NW-side model, support to report metrics for UCI reporting overhead reduction. |
| OPPO [10] | Proposal 13: For UE-side model, additionally support Type 2 (indication/request/report from UE to gNB) performance monitoring.  Proposal 14: For UE-side model, support the beam prediction accuracy (Alt.1) and probability/confidence of model output (Alt.3) as performance metrics.  Proposal 15: For Type 1 Option 1 and Type 2 performance monitoring, discuss and specify (if necessary) the LCM-related events.  Proposal 16: For performance monitoring, discuss and specify (if needed) LCM events based on beam prediction accuracy (Alt.1) and probability/confidence (Alt.3) within a time window. |
| Fujitsu [11] | Proposal 13:   * Regarding performance monitoring for BM Case-1 with UE side model, Type-2 performance monitoring should also be supported.   Proposal 14:   * Regarding Option 2 of Type-1 performance monitoring (UE assisted monitoring) for BM Case-1 with UE side model, RAN1 to further discuss the performance metric, and the following alternatives are preferred.   + 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 15:   * Regarding Option 2 of Type-1 performance monitoring (UE assisted monitoring) for BM Case-1 with UE side model, RAN1 to further discuss the reference signal configuration and reporting enhancement.   Proposal 16:   * Regarding Type-2 performance monitoring (UE-side monitoring) for BM Case-1 with UE side model, Rel-16 SCell BFR framework could be used as starting point. RAN1 to further discuss:   + the timer/counter/threshold configuration for performance monitoring   + how to report the information that the current functionality doesn’t work well and the newly activated functionality.   Proposal 17:   * Regarding Type-2 performance monitoring (UE-side monitoring) for BM Case-1 with UE side model, RAN1 to further discuss the corresponding reference signal configuration.   Proposal 18:   * Regarding Type-2 performance monitoring (UE-side monitoring) for BM Case-1 with UE side model, RAN1 to further discuss the performance metric, and the following alternatives are preferred.   + 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 20:   * Regarding NW-side monitoring for BM Case-1 with NW-side model, RAN1 to further discuss the performance metric, and the following alternative is preferred.   + Alt.4: The L1-RSRP difference evaluated by comparing measured RSRP and predicted RSRP.   Proposal 21:   * Regarding NW-side monitoring for BM Case-1 with NW-side model, RAN1 to further discuss the reference signal configuration and possible reporting enhancement, e.g., quantization of L1-RSRP. The high-resolution quantization and non-differential RSRP could be considered for ground truth data for performance monitoring.   Proposal 26:   * Regarding performance monitoring for BM Case-2 with UE side model, RAN1 to consider similar design scheme as BM Case-1 with UE side model.   Proposal 30:   * Regarding the performance monitoring for BM Case-2 with NW-side model, similar scheme as BM Case-1 with NW-side model could be considered.   Proposal 31:   * RAN1 to discuss LCM operation which is specific to beam management use case. |
| CATT [12] | Proposal 13: For performance monitoring of BM-Case1 and BM-Case2, the following performance metrics can be supported：   * Alt.1: Beam prediction accuracy related KPIs, including Top-1 beam prediction accuracy, Top-K/1 beam prediction accuracy, Top-1/K beam prediction accuracy and Top-1 beam prediction accuracy within 1 dB margin; * Alt.4: The L1-RSRP difference evaluated by comparing measured RSRP and predicted RSRP, including the L1-RSRP difference of beams in Set B and L1-RSRP difference of predicted Top-K beam(s).   Proposal 14: For Type 1 performance monitoring of Option 1, support per-sample based reporting as the baseline.  Proposal 15: For Type 1 performance monitoring of Option 2, support both per-sample based and multiple-sample based reporting with considering the following aspects:   * the number of samples within a monitoring window, e.g., configured by the gNB; * the report contents, e.g., the statistic of the performance metric; * the specific events and uplink resources for event-based reporting.   Proposal 16: For Type 2 performance monitoring for UE-sided model, the request signaling for performance monitoring should indicate the information of the AI/ML functionality and the performance metric.  Proposal 23: For performance monitoring of NW-sided model, the enhancements of report contents can be considered, where the report contents are relevant to the performance metric and can be configured by the network.  Proposal 24: For performance monitoring of NW-sided model, the rule for UE selecting the reported beam(s) can be specified to ensure the NW can obtain accurate performance metrics. |
| Lenovo [14] | Proposal 17: 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 18: 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 19: 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.   Proposal 20: For a monitoring sample, the association of the beam measurement for Set B and the beam measurement for benchmark/reference should be ensured.  Proposal 21: Support event triggered beam report for hybrid performance monitoring for UE-side AI/ML model |
| Sony [15] | Proposal 4 : Support for defining event(s) to trigger reporting for monitoring, such as the difference between predicted results and measured results at the inference stage in BM-Case 1.  Proposal 5 : L1-RSRP difference between the predicted and actual values should serve as the performance metric when the model output includes predicted L1-RSRP.  Proposal 6 : Model failure detection mechanism can be defined according to the process of the beam failure detection. |
| InterDigital [16] | Proposal 19: For AIML LCM in BM, consider procedures for identification of need of AI/ML model recovery, UE request/gNB trigger and AI/ML model recovery.  Proposal 20: Support Option 2 (UE-assisted performance monitoring).  Proposal 21: Support a UE computed performance metric (e.g., UE’s local beam prediction accuracy), calculated based on measured and predicted beam qualities of Top-1/Top-K beam(s). |
| Panasonic [17] | Proposal 10: For performance monitoring of UE-sided model, we propose to take performance metric that is more directly related system performance like Alt. 2 (link quality related KPIs) or Alt. 4 (L1-RSRP difference evaluated by comparing measured RSRP and predicted RSRP). Then, if necessary, as an intermediate KPI, Alt. 1 (beam prediction accuracy related KPIs) or Alt. 3 (performance metric based on input/output data distribution of AI/ML) can be taken together with Alt. 2 or Alt. 4.  Proposal 11: Group-based beam reporting can be enhanced to support performance monitoring for NW-sided model. |
| Nokia [19] | Proposal 7: For BM-Case1 and BM-Case2, considering NW-sided performance monitoring for beam prediction related CSI reporting, discuss the following variants,   * Case1: A different CSI report is used to support NW-sided performance monitoring.   + RS resource set: The NW can use a different CSI report to get beam measurements/reporting for a monitoring RS resource set (as preferred by the NW) within the legacy CSI reporting framework.   + L1-RSRP and RS index of Top-M beams of monitoring RS set is supported as the content of the report   + Note: Spec impact may only expect if M is defined separately for monitoring purpose. * Case 2: The same CSI reporting configuration is used for both monitoring and inference.   + Option 1: Consider monitoring RS resource set = Set A (same RS resource set for inference and monitoring).   + Option 2: Monitoring RS resource set is configured/indicated separately from Set A.   + Option 3: Monitoring RS resource set is determined by the UE based on active TCI states or inference outcome(s).   + For Options 1-3, the NW configures (associated to the CSI report) the reporting timelines and reporting quantities for the monitoring RS resource set.     - L1-RSRP and RS index of Top-M beams of monitoring RS set is supported as the content of the report   Proposal 8: For BM-Case1 and BM-Case2, considering UE-assisted performance monitoring for a beam prediction related CSI reporting, discuss whether the following options can be used for the reporting of performance monitoring related KPIs,   * Option 1: reporting of “Top-K beam prediction accuracy” corresponding to predicted Top-K beam IDs, where K is configurable to the UE. * Option 2: reporting of “L1-RSRP difference” corresponding to Top-1 predicted beam. * Option 3: reporting of “L1-RSRP difference predicted” corresponding to predicted L1-RSRP of Top-1 predicted beam, if predicted L1-RSRP is supported by AI/ML model output. * Note: The UE shall refer to the KPIs definition contained in TR 38.843, more realistic assumptions for ideal L1-RSRP and genie-aided beam shall be described if any of the above metrics are defined in the normative specifications.   Proposal 9: For BM-Case1 and BM-Case2, considering UE-assisted performance monitoring for a beam prediction related CSI reporting, study whether the same CSI report (CSI report that used for inference) or a different CSI report should be used.   * For the case where the same CSI report is used for monitoring and inference. Discuss following variants,   + Option 1: Consider monitoring RS resource set = Set A (same RS resource set for inference and monitoring).   + Option 2: Monitoring RS resource set is configured/indicated separately from Set A.   + Option 3: Monitoring RS resource set is determined by the UE based on active TCI states or inference outcome(s).   + For Options 1-3, the NW configures (associated to the CSI report) the reporting timelines and monitoring KPIs for the monitoring RS resource set. * For the case where different CSI reports are used for monitoring and inference, NW can configure/indicate the monitoring RS resource set (resourcesForChannelMeasurement) within the legacy CSI reporting framework.   Proposal 10: For BM-Case1 and BM-Case2, considering UE-assisted performance monitoring for beam prediction related CSI reporting, discuss the following for event-based reporting,   * Details of monitoring RS resource(s) and details of RS resources for prediction. * Define details of failure events and discuss following variants,   + Event-1: Predicted beam accuracy of the set of predicted beams being below a threshold accuracy.   + Event-2: Predicted L1-RSRP of the set of predicted beams being below a threshold value.   + Event-3: Hypothetical BLER of a predicted beam of the set of predicted beams being below a threshold. * Details of the events associated configurations including definition of thresholds, counters and timers configured to determine the failure instances for a beam prediction related CSI report. * Details of reporting for failure event, including reporting content. * Strive to use similar mechanisms as in BFR procedures.   Proposal 11: For BM-Case1 and BM-Case2, considering UE simultaneously supports both Option 1 (NW-sided performance monitoring) and Option 2 (UE-assisted performance monitoring), study/discuss whether a UE indicates its preference for which option and how to support it.  Proposal 12: RAN1 to prioritize work on specifying NW-side performance monitoring and discuss/study different options for UE-assisted performance monitoring. Deprioritize UE-side performance monitoring.  Proposal 17: There is no additional requirement for defining any new signalling for functionality LCM for beam prediction use case. |
| Ruijie [20] | Proposal 1: For Option 1 in Type 1 performance monitoring, it is proposed that beam information related to predicted Top 1 or Top K beams and/or ground truth of the target Set A resources should be also included in reporting contents.  Proposal 2: For Option 2 in Type 1 performance monitoring, UE should directly report performance monitoring decisions/results (e.g., beam prediction results/performance/accuracy) or events based on its calculated performance matric(s).  Proposal 3: For Option 2 in Type 1 performance monitoring, in order to reduce spec effort, it is proposed to use existing CSI framework for reporting, at least as a starting point.  Proposal 4: Support event(s)-based reporting for Option 1 and/or Option 2. |
| Samsung [21] | Proposal 20. For the option 1 of Type 1 performance monitoring, the existing CSI reporting mechanism and the CSI reporting mechanism for UE-side model inference are used as baseline.   * FFS: whether to use single CSI-ReportConfig for the report for both prediction results and measurement results for Set A.   Proposal 21. For the option 2 of Type 1 performance monitoring, at least for BM-Case1, consider UE to indicate the calculation result of the following performance metrics:   * Opt 1. The measured Top-K beam(s) of Set A and the predicted Top-K beam(s) of Set A are all the same or not. * Opt 2. The L1-RSRP difference between the measured Top-K beam(s) of Set A and predicted Top-K beam(s) of Set A are larger than a threshold value or not. * Opt 3. The probability information of Top-1 beam of Set A is lower than a threshold value or not.   Proposal 22. For the option 2 of Type 1 performance monitoring, at least for BM-Case1, consider the following events to trigger UE reporting/notification:   * Event-1: The measured Top-K beam(s) of Set A and the predicted Top-K beam(s) of Set A are different * Event-2: The L1-RSRP difference between the measured Top-K beam(s) of Set A and predicted Top-K beam(s) of Set A are larger than a threshold value * Event-3: The probability information of Top-1 beam of Set A is lower than a threshold * FFS: The content of the reporting/notification * FFS: The configuration of Set A and or Set B   Proposal 23. For UE-side AI/ML model, support Type 2 performance monitoring.  Proposal 24. For the support of Type 2 performance monitoring of UE-side AI/ML model, consider the extension of the CSI reporting mechanism, e.g., CSI-ReportConfig with reportQuantity set to ‘none’. |
| Transsion [22] | Proposal 7: Regarding performance metric(s) of BM-Case1 and BM-Case2, support the following alternatives:   * 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 8:For performance monitoring of NW-side model, configuration of Set B and/or Set A and the reporting of measurement results of Set B and/or Set A need to be specified.  Proposal 9: For Option 2 of Type 1 performance monitoring, consider the following alternatives can be considered:   * Alt-1: Quantization of performance metric; * Alt-2: The statistic values of the performance metric; * Alt-3: A specific event. |
| ETRI [23] | Proposal 4: Support beam prediction accuracy and L1-RSRP difference as performance metrics for UE-assisted performance monitoring.  Proposal 5: For UE-assisted performance monitoring, support the following event to report to the NW when the counter exceeds a certain threshold.   * The Counter is incremented if the predicted Top K beam results differ from the actual measured results and/or if the L1-RSRP difference exceeds a predefined value.   Proposal 6: For Type 2 performance monitoring, the UE can transmit information regarding the activation/deactivation, switching, or fallback of the functionality.  Proposal 7: Support Monitoring IDs to differentiate monitoring processes for multiple UE-sided models.  Proposal 8: For the NW-sided model, support an additional beam set for performance monitoring. |
| CAICT [24] | Proposal 2: For option 2 of type 1 UE-sided model monitoring, report probability information of the predicted beam of Top 1 or Top K beams（Option D）is proposed.  Proposal 3: Event based monitoring is not proposed for UE-sided model monitoring for Type 1 monitoring option 1/2.  Proposal 4: UE initiated fall back operation should be considered as a special case for type 2 performance monitoring. |
| DOCOMO [25] | Proposal 14: Not support type 2 performance monitoring.  Proposal 15: Support the following performance metric for reporting beam information and/or RSRP of predicted top K beam(s) (Opt1 and Opt2).  ・Beam prediction accuracy related KPI (for Opt1/2).  ・L1-RSRP difference evaluated by comparing measured RSRP and predicted RSRP (for Opt2).  Proposal 16: Support reporting of performance metric values in UE-assisted performance monitoring.  Proposal 17: Support the following triggering mechanism of UE-assisted performance monitoring.   * Based on NW configuration/indication * When performance metric satisfies some conditions, such as larger/lower than thresholds |
| Sharp [26] | Propsoal1: Regarding the performance metrics of model monitoring, support Beam prediction accuracy related information. The beam prediction accuracy can be defined as Beam Prediction Accuracy= TT⁄((TT+FT)), where TT (True Top-K beams) denotes that a beam is predicted as one of the Top-K beams and is actually one of the measured Top-K beams, and FT (False Top-K beams) denotes that a beam is predicted as one of the Top-K beams but the beam is not actually one of the measured Top-K beams. Thus, the beam prediction accuracy ranges from 0 to 1.  Proposal 2: For AI/ML model performance monitoring for BM-Case1 and BM-Case2, support the following Alt.1 as the benchmark/reference 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).   Proposal 3: For UE-side AI/ML model, additional contents beyond the measurement results is unnecessary for option 1 of Type 1 performance monitoring.  Proposal 4: For UE-side AI/ML model, in option 2 of Type 1 performance monitoring, support event-driven performance metric reporting.  Event: a calculated beam prediction accuracy is lower than a threshold value. |
| Qualcomm [27] | Proposal 10 For UE-side beam prediction, when the performance monitoring set is equal to Set A for a set of performance monitoring instances, study the following metric for performance monitoring:   * Definition for Top-K beam prediction accuracy with L1-RSRP margin:   + Define as the total number of performance monitoring instances.   + Define as the number of performance monitoring instances (out of ) for which the following statement holds:     - The highest measured L1-RSRP of Top-K predicted beams is within a margin of measured L1-RSRP of best measured beam ID from performance monitoring set   + Top-K beam prediction accuracy with L1-RSRP margin is defined as the ratio . * Note 1: “performance monitoring set” is the set of RSs that are to be measured for performance monitoring, per performance monitoring instance. * Note 2: The above metric is at least applicable to Type 1 performance monitoring, Option 2 (UE-assisted performance monitoring).  Proposal 11 For UE-side beam prediction, for a given performance monitoring instance in which the performance monitoring set is equal to Set A, study the following metric for performance monitoring:   * The L1-RSRP difference between measured L1-RSRP of Top-1 predicted beam ID from Set A and measured L1-RSRP of best measured beam ID from Set A, per performance monitoring instance. * Note 1: “performance monitoring set” is the set of RSs that are to be measured for performance monitoring, per performance monitoring instance. * Note 2: The above metric is at least applicable to Type 1 performance monitoring, Option 2 (UE-assisted performance monitoring).  Proposal 12 For UE-side beam prediction, study details of performance monitoring metrics based on availability of RS for performance monitoring (in particular when only a subset of Set A is measured for performance monitoring). Proposal 13 For UE-side beam prediction, and for UE-assisted performance monitoring, study details of performance monitoring reports (contents, frequency of report, carrier), at least as a function of performance monitoring metrics. Proposal 14 For BM-Case1 and BM-Case2 with a UE-side AI/ML model, and for Type 1 performance monitoring, support event-triggered performance monitoring report for both Option 1 and Option 2   * FFS: definition of events |
| Indian Institute of Tech (M), IIT Kanpur [28] | Proposal 3: For performance monitoring of UE-sided models, UE can report RSRP difference between true and predicted value for Top 1 or Top K beams along with confidence information of each RSRP difference.  Proposal 4: For performance monitoring of UE-sided models, UE can report Top K beams with confidence information/probability information.  Proposal 5: To address the report overhead issue, the report quantities, a) RSRP difference of true and predicted Top K beams with confidence information b) Top K beams with probability information can be formulated as Events and can be triggered by UE to report the model performance. |
| Fraunhofer HHI, Fraunhofer IIS [29] | Proposal 5: For monitoring UE-sided models, support 2-phase monitoring with varying frequencies and reporting detail.  Proposal 6: Consider indication-based and event-based switching into a validation phase. Events may be defined based on agreed performance metrics.  Proposal 8: Support a model monitoring configuration that allows for collecting data for model training and monitoring of inactive models. |
| NTU [30] |  |
| ITL [31] | Proposal 21: For performance monitoring of BM-Case1 and BM-Case2, following performance metrics can be supported:   * Alt.1: Beam prediction accuracy related KPIs, including Top-1 beam prediction accuracy, Top-K/1 beam prediction accuracy, Top-1/K beam prediction accuracy and Top-1 beam prediction accuracy within 1 dB margin; * Alt.4: The L1-RSRP difference evaluated by comparing measured RSRP and predicted RSRP, including the L1-RSRP difference of beams in Set B.   Proposal 22: For UE-side model monitoring of Option 2 of Type 1, it can be considered to define the performance metrics representing beam prediction accuracy and related statistical values/types for the reporting.  Proposal 23: It is proposed to support event-triggered UE reporting for UE-sided Type 1 performance monitoring.  Proposal 24: For Type 2 performance monitoring of UE-side model, it is proposed to define new report quantity including no reporting (e.g. ‘none’) and the monitoring decision (e.g. activation, deactivation or fallback) in CSI-ReportConfig. |
| KDDI [32] | Proposal 5: For Type 1 Option 1, reuse the existing CSI framework for reporting measurement results, with minimal enhancements if necessary.  Proposal 6: Prioritize support for Option B (beam prediction accuracy) and Option C (RSRP difference) for UE-assisted performance monitoring.  Proposal 7: For UE-assisted performance monitoring, support reporting of aggregated statistics (e.g., percentage of correct predictions, average and maximum RSRP differences) over a configurable time window for beam prediction accuracy and RSRP difference metrics.  Proposal 8: Further study is needed to clearly define and evaluate the usefulness of Option D (probability information) and Option E (confidence information).  Proposal 9: Deprioritize Option F (hypothetical BLER-like metrics), as it does not directly evaluate the AI/ML model's monitoring performance in beam management tasks. |
| Huawei/HiSi [33] | Proposal 25: Except for the data collection for measurement report, there is no need to specify a monitoring procedure/metric for the NW-side model.  Proposal 26: For the monitoring types of UE-side model monitoring for both BM-Case 1 and BM-Case 2:   * For Type 1 Option 1, the procedure is the same as measurement report, and there is no need to introduce event-based reporting. * For Type 1 Option 2, consider both reporting a calculated metric and reporting an event.   + For metric-based reporting , consider report per sample, report per set of samples, or report of the statistical value.   + For event-based reporting, event is triggered when the prediction accuracy metric is lower than the threshold and satisfies a timer/counter. * For Type 2, UE does not report CSI, but reports the recommended monitoring decision (e.g., fallback) to NW.   + gNB may configure a threshold criterion to facilitate the UE to perform model monitoring.   Observation 5: Model monitoring has more stringent requirements on latency than training, which makes L1 signaling more suitable.  Proposal 27: For the monitoring Type 1 (Option 1 and Option 2) of UE-side model monitoring, consider L1 signaling with higher priority.  Proposal 28: For UE-side model monitoring Type 1 Option 2 and Type 2, the association between the label (or measured Set A) and the predicted CSI (or measured Set B) should be indicated/configured to UE.   * E.g., the associated CSI-ResourceConfigId of Set B and/or the time relationship to the predicted CSI can be indicated for the measurement of Set A.   Proposal 29: For UE-side model monitoring with dedicated measurement of Set A, consider the following types of metrics as a starting point.   * Beam index accuracy, e.g., accuracy between genie-aided best beam index(es) from Set A and predicted best beam index(es). * Measured L1-RSRP gap information, e.g., gap of measured L1-RSRPs between genie-aided best beam index(es) from Set A and predicted best beam index(es).   Proposal 30: For UE-side model monitoring without dedicated measurement of Set A, if needed, consider the following types of metrics:   * Probability information of predicted Top beam(s). * Confidence information calculated from the measured L1-RSRP(s) and predicted RSRP(s) for beams in Set B. |
| Xiaomi [34] | Proposal 5-1: Support following two performance metrics with high priority for performance 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   Proposal 5-2: Both of the following two Benchmark/reference for performance comparison should be supported.   * 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)   Proposal 5-3: For Type 1 performance monitoring of UE-side AI/ML model, both NW-side initiated and UE-side initiated performance monitoring can be supported. NW-side initiated can be based on measurement/report configuration via RRC and UE-side initiated can be based on SR and UL MAC CE with the preferred resource configuration of set B and set A.  Proposal 5-4: The number of the inference instances should be configured to UE for calculation of a statistic value of performance metric for performance monitoring.  Proposal 5-5: If the performance metric is the L1-RSRP difference, not consider the beams in set B.  Proposal 5-6: For UE-side AI/ML model, support Type 2 performance monitoring and it can be initiated by UE-side based on SR and UL MAC CE.  Proposal 5-7: For UE-side AI/ML model with Type 2 performance monitoring, it is better to indicate UE’s decision to NW for consistency of the NW-side additional condition for the new applied UE-side model.  Proposal 5-8: For UE-side AI/ML model with Type 2 performance monitoring, configure an event with a threshold to assist UE to make the decision.  Proposal 5-9: For performance monitoring of network-side AI/ML model, support to report measurement results of set B and set A separately. Set B can be reported based on beam report, and set A can be reported by MAC CE or RRC with multiple samples.  Proposal 5-10: For performance monitoring for network-side AI/ML model, support an event-triggered report if the indicated TCI state is different from the best beams obtained by measurements.  Proposal 5-11: Confirm the necessity of assessment/monitoring of inactive models / functionalities, with the following assumptions as the starting point:   * One way to monitor inactive models/functionalities is by activating them and reusing mechanisms defined for monitoring of active models/functionalities. * The following aspects may be considered for further study or in WI to assess the applicability and expected performance of an inactive model/functionality:   + Configuring an AI/ML model for monitoring without activation (e.g., monitoring-only mode without reporting predicted beams in BM Case 1 and 2)   + Dataset delivery from the network to the UE for assessment/monitoring of the applicability and expected performance of the model/functionality.   + NW may provide performance criteria/preference for UE’s model selection.   + Other aspects are not precluded for further study or specification. |
| Kyocera [35] | **Proposal 14:** For the performance monitoring of a UE side AI/ML model, for the content of the report of type 1 option 1 monitoring, support the following as a starting point:   * Model related outputs:   + Beam information of the top-K predicted beams.   + Predicted RSRP if supported by the AI/ML model.   + Probability information of the predicted beams.   + Confidence information of the of the RSRP if supported by the AI/ML model. * Performance related measurements from the resource set configured for monitoring:   + CRI/SSBRI   + L1-RSRP of the top-K beams   Proposal 15: For type 1, option 2, performance monitoring of a UE side AI/ML model, regarding the FFS of:   * How to report- assess the impact of adopting the following methods:   + Sample based performance metric reporting in order to provide the NW with frequent updates about the AI/ML model performance.   + Statistical based performance metric reporting in order to reduce the UL overhead on the expense of less information being available at the gNB. * What to report- Support using the following metrics as a starting point:   + The error between the predicted RSRP and the measured L1-RSRP, where the predicted RSRP is based on AI/ML output.   + The beam predication accuracy of the top-K beams, where the beam prediction accuracy is the percentage of the beam determined from RS measurements for performance monitoring to be one of the top-K predicted beams by the AI/ML model.   Proposal 16: For UE-side AI/ML performance monitoring, RAN1 should further study the following:  • An event is defined as the scenario when the performance metric falls above (or below) a certain threshold.  • The concept of triggering a report based on specific events is not applicable to option 1. It is, however, relevant to option 2, where the UE calculates performance metrics and can detect events.  Proposal 17: For UE side AI/ML performance monitoring, type 2 monitoring has no specification impact, and it should be left up to the UE vendor implementation.  Proposal 18: For a NW side AI/ML model performance monitoring, there is no specification impact is needed and it can be left up to the NW vendor implementation. |
| LGE [36] | Observation #2: For Option 1 (NW-side performance monitoring) of Type 1 performance monitoring, there may be no specification impact, e.g., NW can exploit more than 4 beam related information report in L1 signaling for performance monitoring purpose.  Proposal #16: For Option 2 (UE-assisted performance monitoring) of Type 1 performance monitoring, support following metrics considering different type of UE-sided AI/ML model:   * Top-K/1 beam prediction accuracy * Difference between measured L1-RSRP and predicted RSRP   Proposal #17: Support event-triggered UE reporting for UE-sided AI/ML performance monitoring.   * Further consider UE report via UCI or SR to request change of Set A configuration, fallback to legacy beam report, holding the report for a while, etc. |
| NVIDIA [37] | Proposal 5: For AI/ML based beam prediction in spatial/time domain, introduce specification support for assistance signalling and procedure for model configuration, model activation/deactivation, model recovery/termination, and model selection.  Proposal 6: For AI/ML based beam prediction in spatial/time domain, introduce specification support for assistance signalling and procedure for model performance monitoring and model update/tuning. |
| NEC [38] | Proposal 20: Support UE to initiate performance monitoring of current AI/ML model if beam failure instance indication or out-of-sync indication is received.  Proposal 21: Support UE to activate candidate models and to initiate performance monitoring of candidate models before model switching or model selection.  Proposal 22: Study simultaneous performance monitoring for multiple candidate models, including how to inform the NW the inactive candidate models and how to request resources and configurations for performance monitoring of the inactive candidate models.  Proposal 23: At least for BM-Case2 performance monitoring, study the method to configure the associated measurement and report resources for obtaining the ground truth in future time instances, and the method to configure the associated measurement and report resources for obtaining the historical measurement results as model input.  Proposal 24: Support L1 signalling for performance monitoring. For Type 1 performance monitoring of UE side model, use the existing CSI framework as a baseline and explore the enhancement for Option 2.  Proposal 25: It is necessary to **consider** selecting a subset of Set A as the monitoring RS resource set.  Proposal 26: For performance monitoring of UE-side model, support to assess the performance for multiple Set Bs to balance beam measurement overhead and performance metrics.  Proposal 27: For Option-2 (UE-assisted performance monitoring), following methodology is used.  − Step-0: gNB configures UE for performance monitoring and reporting (FFS whether gNB configures the performance metric)  − Step-1: UE determines the performance metric based on the received configuration  − Step-2: UE reports the results of performance metric to the gNB based on the reporting configuration  − Step-3: gNB decides what model management decision to take based on the report received from UE.  Proposal 28: For Option-2 (UE-assisted performance monitoring), support UE to determine and report the result of following performance metrics.  − Top-K/1 (%): the percentage of "the Top-1 measured beam is one of the Top-K predicted beams  − Top-1/K (%): the percentage of "the Top-1 predicted beam is one of the Top-K measured beams"  Proposal 29: Support UE to report the following performance metrics:  − The L1-RSRP difference between the measured L1-RSRP and predicted L1-RSRP of beam(s) in Set A.  Proposal 30: Support UE to report the following performance metrics:  − Probability information of the predicted beam of Top 1 or Top K beams.  − Confidence information of the predicted RSRPs.  Proposal 31: Support UE to report probability(ies) of predicted Top K beam(s) based on some pre-defined interval or threshold/criterion. Otherwise, if the probability is not reported, the probability should be used as one of the determining factors when reporting predicted beams at least for classification model.  Proposal 32: The confidence information should be defined as a confidence interval or prediction interval associated with predicted L1-RSRPs at a specific confidence level (e.g., 95%).  **Proposal 33:** Monitoring based on data distribution should be supported. |
| MTK[39] | Proposal 7: For performance metrics for monitoring an AI/ML beam management model, RAN1 focus on the following metrics calculated at UE and/or gNB side:   * Beam prediction accuracy:   + Statistical results on full/subset of Set A measurements * Beam prediction ranking/ordering accuracy   + Note: post-processing of the probability/confidence information   + One shot result on full/subset of Set A measurements   + Statistical results on full/subset of Set A measurements * L1-RSRP difference:   + The L1-RSRP difference between the measured L1-RSRP of the predicted beam and the best L1-RSRP in full/subset of Set A * Predicted RSRP difference:   + The RSRP difference between the measured L1-RSRP and predicted RSRP of a set of beams   + The RSRP difference between the measured L1-RSRP of current beam and the predicted RSRP of the predicted Top 1 beam   Proposal 8: For Type 1, option2, UE-assisted performance monitoring, consider the following conditions to trigger a UE report:   1. A report triggered/activated by NW 2. UE initiated performance monitoring reporting when an event happens (FFS on event)   Proposal 9: For Type 1, option2, UE-assisted performance monitoring, consider the following two alternatives of the reported content:   1. UE reports the performance metric(s) 2. UE reports the performance metric(s) and LCM decisions (model switching/activating/deactivating request)   Proposal 10: For Type 1, option2, UE-assisted performance monitoring, to facilitate UE to detect a monitoring event for reporting, considering NW signaling to UE the following aspects to define an event,   * 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 |
| KT [41] | Proposal 5. Support L1-RSRP difference evaluated by comparing measured RSRP and predicted RSRP as a performance metric.  R1-240xxxx AI in BM summary-v008\_ETRI\_Xiaomi.docx |
| Meta [21] | Proposal 5: For UE side model with monitoring Type-1, Option-1, measurement report with L1-RSRP and RS index is sufficient and other contents may not be needed. Consider configuration of monitoring resource set and time window or filtering based reporting for additional reliability of performance monitoring.  Proposal 6: For UE sided model with monitoring Type-1, Option-2, beam prediction accuracy and RSRP difference reporting may be supported as configurable reporting contents but confidence information and/or probability information may be optional based on UE model capabilities.  Proposal 7: For reporting RSRP difference information, the baseline can be for the current indicated beam where UE can measure and predict the RSRP. For Top-1/K beams which are predicted but not in set B, further discuss how to evaluate RSRP difference  Proposal 8: For UE sided model with monitoring Type-1, Option-2, consider supporting UE report of fallback to non-AI/ML methods  Proposal 9: Support Type 2 performance monitoring of UE sided models with reporting for AI/ML model switching/activation and indication of non-AI/ML fallback  Proposal 10: For UE sided model with Type-1 performance monitoring, support event triggered monitoring where the gNB configures one or more events to the UE  Proposal 11: For UE sided model with Type-1 performance monitoring, support at least Events 1,2 and 3  Proposal 12: For UE sided model with Type-2 performance monitoring, support event driven indication of fallback or model switching/activation.  Proposal 13: Consider UE assisted performance monitoring for NW sided models. |

### Issue #2.1: For NW sided model and/or UE sided model, signaling for performance monitoring

##### (FL0) Proposal 2.1:

For BM-Case1 and BM-Case2 with a UE-side AI/ML model, for Type 1 performance monitoring Option 1 (NW-side performance monitoring), L1 signalling is used to send the measurement results to NW for the calculation of performance metrics at NW.

* Note: this does not preclude to use higher layer signalling if other contents are introduced for Type 1 performance monitoring Option 1.

|  |  |
| --- | --- |
| Supported companies | FL |
| Companies | View |
| FL | It is hard to extend L1 for Option 2 without knowing what to report |
| OPPO | Support the FL proposal.  Since NW is to calculate the performance metric(s) based on raw UE reporting in Option 1, we tend to think the reporting content would be quite similar to existing beam reporting. Hence, L1 signaling can be reused as much as possible. |
| HW/HiSi | Agree |
| Fujitsu | Generally fine with the FL proposal. |
| New H3C | OK in general |
| TCL | Agree |
| ETRI | We support FL’s proposal |
| Xiaomi | If the metric needs to be calculated by measurement results of multiple time instance, we think higher layer signaling can be used to report the measurement results of multiple time instance in one report. We suggest the following update:   * Note: this does not preclude to use higher layer signalling ~~if other contents are introduced~~ for Type 1 performance monitoring Option 1. |
| ZTE | Support. There's a typo in the main bullet where 'UE-side' should be 'UE-sided'. |
| Qualcomm | We agree with the fact that L1 signaling can be used for this purpose, but in our view, it is important to highlight the fact that we can save on L1 signaling if we detect an event that makes reporting to network unnecessary. Please note the following agreement from RAN1 #117:  Agreement (RAN1 #117)  For BM-Case1 and BM-Case2 with a UE-side AI/ML model:   * Support Type 1 performance monitoring, including the following two options:   + Option 1 (NW-side performance monitoring):     - UE sends a report to NW (for the calculation of performance metric at NW)       * Measurement results from resource set for monitoring, e.g., L1-RSRP and/or RS index is supported as the content of the report       * FFS on other contents     - The report is at least configured/triggered by NW     - Note: this may or may not have additional spec impact   + Option 2 (UE-assisted performance monitoring):     - UE calculates performance metric(s)       * FFS how to report and what to report   + FFS whether to trigger the report based on event(s) for Option 1 and/or Option 2 * FFS Type 2 performance monitoring   Given the FFS above and as an example, for a given performance monitoring instance, if UE can do a quick check and figure out that Top-1 beam is actually among the Top-K predicted beams, then what is the justification for sending the L1-RSRPs as well as beam IDs to the NW for comparison, in L1 signaling? For that particular instance, the L1 report can be skipped altogether. So, in our view, L1 signaling would be a reasonable choice if it is event-based and can be skipped if there is no need for reporting. With that said, we suggest the following:  Updated Proposal 2.1  For BM-Case1 and BM-Case2 with a UE-side AI/ML model, for Type 1 performance monitoring Option 1 (NW-side performance monitoring), L1 signalling is used to send the measurement results to NW for the calculation of performance metrics at NW.   * Note 1: this does not preclude to use higher layer signalling if other contents are introduced for Type 1 performance monitoring Option 1. * Note 2: Event-based L1-signaling can be used to avoid sending L1 report when not needed.   FFS: definition of events |
| InterDigital | Fine |
| Ericsson | Ok in general. Unclear if this agreement is needed. The NW should be able to reuse existing reporting (given there is a mechanism for the UE to report inference results). |
| LG | L1 signaling is fine. We think that already agreed more than 4 beam related reporting for NW-side AI/ML can be exploited for performance monitoring, too.  Regarding Qualcomm’s comment regarding event-based L1-signaling for Type 1 - Option 1 performance monitoring, I think that event-based L1 reporting from UE side for Type 1 - Option 1 performance monitoring is not appropriate since NW calculates metrics. |
| CATT | We support using L1 signaling for Option 1 Type 1. First, there is limited spec impact to support L1 signaling for Option 1 type1. Secondly, for network, these raw data can used not only for performance monitoring, but also other purpose, such as measurement or beam management. It’s nature to support 1 using L1 signaling for Option 1 type1. |
| vivo | Support |
| Fraunhofer | Agree. |
| Nokia | This looks OK as a clarification to the earlier agreement on performance monitoring. |
| KDDI | Support |
| Panasonic | Fine with proposal. |
| MediaTek | Whether L1 or higher layer signaling is used should depend on the content of the measurement results in a report. If the content is beam ID + L1-RSRP of the best beam(s) in the set of measurement, we agree that L1 signaling can be used. But if the content is L1-RSRP for all the beams in Set A (for calculating RSRP difference between the measured and predicted), we don’t think L1 signaling should be used. It will create a huge burden on UE’s reporting. |
| Spreadtrum | Agree. The report content seems the same as current measurement report. Thus L1-signaling is the easy way. |
| CMCC | Ok. |
| Sony | Support, when NW calculates the performance metrics necessary for UE to report a measurement report. |
| Sharp | Support. |
| Futurewei | Support FL’s proposal. |
| NEC | Support |
| Lenovo | Fine |
| Kyocera | We agree with the proposal |

##### (FL1) Proposal 2.1a:

For BM-Case1 and BM-Case2 with a UE-side AI/ML model, for Type 1 performance monitoring Option 1 (NW-side performance monitoring), L1 signalling can be used to send the measurement results to NW for the calculation of performance metrics at NW.

* Note: this does not preclude to use higher layer signalling ~~if other contents are introduced~~ for Type 1 performance monitoring Option 1.
* Note 2: This does not preclude to introduce event to trigger the L1 report
* Note 3: measurement results refer to “measurement results from resource set for monitoring, e.g., L1-RSRP and/or RS index is supported as the content of the report in previous agreement”

|  |  |
| --- | --- |
| Companies | View |
| FL |  |
| MediaTek | When we say “L1 signalling can be used”, does it mean there is no spec impact, or we will discuss the impact later? Since the definition of “measurement results” is not clear in this proposal, we think it is hard to conclude that “L1 signalling can be used” without any spec impact. For example, if the measurement results is the L1-RSRP of all the beams in Set A, if such beam report is configured periodically, how to use the limited PUCCH resource for transmitting such large size of a beam report? |
| FL | Please express your strong concern, if any |
| Kyocera | In general, we are fine with the FL proposal. One comment:  It is unclear for us the defination of an event in case of type 1, option 1 moniroting. As per our understanding, option 1 is a NW side monitoring where the UE reports the ground truth measurements and the ML model outputs consistently (as long as it has any) with no further processing or calculation of any performance metric. Given this, it is unclear for us how an event could be defined in this case.  We propose a slight modification for the proposal as follows:  For BM-Case1 and BM-Case2 with a UE-side AI/ML model, for Type 1 performance monitoring Option 1 (NW-side performance monitoring), L1 signalling can be used to send the measurement results to NW for the calculation of performance metrics at NW.   * Note: this does not preclude to use higher layer signalling ~~if other contents are introduced~~ for Type 1 performance monitoring Option 1. * ~~Note 2: This does not preclude to introduce event to trigger the L1 report~~   Note 3: measurement results refer to “measurement results from resource set for monitoring, e.g., L1-RSRP and/or RS index is supported as the content of the report in previous agreement” |
| New H3C | OK in general |
| CMCC | OK in general. |

### Issue #2.2: For UE sided model, how to report and what to report for Type 1 Option 2

Summary of position:

* Option B: Report the beam prediction accuracy related information based on measurements [of a window]
  + FL: One shot is not reliable. I think in a window should be considered
  + Supported by CMCC, Ericsson, vivo, OPPO, NOKIA, Huawei, KDDI, ITL, ETRI,,ruijie, Interdigital, DoCoMo, Qualcomm? Samsung?(Same or not), xiaomi, Kyocera,LGE,NEC,MTK,meta
* Option C1: Report the RSRP difference information between the measured and predicted
  + Supported by Tejas, Ericsson (statistical prediction performance metrics (e.g., mean, 10th, 50th, 90th percentile), vivo, NOKIA, KDDI,, ITL(of Set B), ETRI, ruijie, Sony, DoCoMo, Qualcomm, Samsung (within a range or not), Xiaomi, Kyocera, LGE ,NEC,MTK,KT,meta
* Option C2: Report the RSRP difference information between the measured L1-RSRP of the Top-1 predicted beam and largest L1-RSRP over Set A.

Supported by Nokia,Huawei,MTK

* Option D: Report probability information of the predicted beam of Top 1 or Top K beams
  + FFS on probability information and the quantization, including
    - #1: The probability information of predicted Top 1
      * Note: this can be treated as report of inference result as well
    - #2: The probability information of each or sum of predicted Top Top-K beams.
      * Note: this can be treated as report of inference result as well
    - #3: Beam information that the probability information comparing to a threshold.
  + Supported by Tejas, CMCC, OPPO, Huawei(w.o SetA, if needed), CAICT, Samsung(#3), Kyocera, NEC, meta(UE capability)
* Option E: Report confidence information of the predicted RSRP error
  + FFS on confidence information and the quantization, including
    - #1: The confidence information of predicted RSRP of the predicted Top 1
      * Note: this can be treated as report of inference result as well
    - #1: The averaged confidence information of predicted RSRP of the predicted Top Top-K beams.
      * Note: this can be treated as report of inference result as well
    - #3: The averaged confidence information of predicted RSRP of all beams in Set A.
    - FL: Only from my personal understanding, proponent shall give clear definition, otherwise, this will be deleted from candidates
  + Supported by Huawei(wo Set A, if needed), Kyocera, NEC, meta(UE capability)
* Option F: Report hypothetical BLER-like metrics
  + FL: if no detailed definition, this will be deleted from candidates
  + Deprioritize: KDDI
* Any update for BM Case 2?

##### (FL0) Proposal 2.2:

For BM-Case1 and BM-Case2 with a UE-side AI/ML model, for Option 2 (UE-assisted performance monitoring), further study the performance metric(s), as well as the corresponding report content(s) (e.g., the metric(s) itself or event(s) based on the metric(s)) and procedure(s), including:

* Option 1: The beam prediction accuracy related information (e.g., Top 1 or Top K prediction accuracy)
  + FFS: percentage based on predictions/measurements in a window or whether or not correct according to one shot prediction/measurement
  + FFS on other details including how to configure the resources for measurements
* Option 2: The RSRP difference information between the measured L1-RSRP and predicted RSRP
  + FFS on statistical prediction performance metrics in a window or according to one shot
  + FFS on the beam(s)/a beam set associated with the RSRP difference information
    - E.g., RSRP difference information between predicted RSRP and measured L1-RSRP of the Top 1 or Top K predicted beam(s)
    - E.g., RSRP difference information between the measured L1-RSRP of the Top-1 predicted beam and largest L1-RSRP over a beam Set (e.g. Set A or a set for monitoring).
* Option 3: The The probability of the predicted beam(s) to be the Top 1 or Top K beam
* Other options are not precluded.

|  |  |
| --- | --- |
| Supported companies | FL |
| Companies | View |
| FL | There are two things, one is what is the metic(s), the other is what to report (the metric itself or an event (comparing to a threshold)) (trigger event is another aspects)  Let’s focus on the metric itself, then further study the report content (e.g., absolute RSRP difference, or whether the RSRP difference is higher than a threshold, etc). Detail on report contents is open to study. Don’t bother that.  Other options have limited supporters. |
| NTT DOCOMO | We do not see the difference between Option 1 and Option 3. If Option 1 and Option 3 correspond to actually measured accuracy and predicted accuracy respectively, Option 3 can be deprioritized due to the uncertainty of prediction accuracy in Option 3. |
| OPPO | We think it is time to decide the performance metric(s) first.  We are fine to study all 3 options.  Perhaps it would be safer to note down that event-triggering based reporting for Type 1 Option 2 is a separate discussion.  On Option 3, we do see the difference with Option 1. The probability output of a model can somehow reflect how confident a predicted beam can be the Top-1 or Top-K beams from model perspective. We believe this information can help NW to make LCM decision. |
| HW/HiSi | The proposal goes into a good direction, some comments for clarity below.   * It seems the intention of this proposal is to discuss the metrics and then whether to directly report the plain metric or to report the metric based on an event is going to be separate discussion. In our understanding, that intention is not totally clear from the main bullet and we suggest to modify accordingly. Therefore, we removed “e.g. metric itself or event based on the metric”, hope that this understanding from our side is correct ☺ * Similar to NW-side performance monitoring, also UE assisted performing monitoring (Option 2) should be performed with L1 signaling due to the time critical aspects of monitoring, at least the L1 signaling should be studied with higher priority. We suggest to express this in this proposal. * We could further clarify on the different types of reporting methods, e.g. per sample, set of samples or statistical value. These reporting mechanisms can apply to all of the suggested options in the proposal, and could be expressed as a separate bullet. There is already some related information in the FFSs under the different options, this could then be replaced by a common bullet. We suggest an update. * The FFS under Option 1 how to configure would apply to all listed options in our understanding. We suggest to make this a main bullet instead. Additionally, examples how to achieve this could already be listed for further guidance of the study. One possibility is to enable the UE for measurement of Set A, the association between the CSI-RS resource sets between Set A (to derive the measurement of the label) and Set B (to derive the predicted CSI) should be indicated/configured to the UE. This could be done by configurations of separate CSI reports for Set A and Set B, and indicating the associated CSI-ResourceConfigId of Set B in the CSI-reportConfig of Set A, which is similar to associating the CSI-ResourceConfigId of Set A in the CSI-reportConfig of Set B for the inference case. Alternatively, the time relationship between resource set(s) of Set A and the predicted CSI to be monitored may also need to be considered in case they are not overlapped in time domain. * For Option 2, another example is to use the gap between L1-RSRP the predicted beam index(es) and the genie-aided best beams of Set A index(es) (or in general of an indicated set, in case not the hole Set A is measured to obtain the global best label).We suggest to add this options as a further example. We suggest to add this is option 2A. * For Option 3, the purpose seems to signal some kind of quality information from the UE to the NW, to let the NW know about the quality of the prediction results. For the regression model, it cannot directly output the RSRP confidence information, but can calculate the confidence information of predicted RSRPs for beams in Set B by comparing with the corresponding measured L1-RSRPs of Set B, so it does not need a dedicated measurement of Set A either. For an equally fair treatment of classification model and regression model, we suggest to add confidence information.  Updated: (FL0) Proposal 2.2: For BM-Case1 and BM-Case2 with a UE-side AI/ML model, for Option 2 (UE-assisted performance monitoring), further study the performance metric(s), as well as the corresponding L1 report content(s) ~~(e.g., the metric(s) itself or event(s) based on the metric(s))~~ and procedure(s), including:   * Option 1: The beam prediction accuracy related information (e.g., Top 1 or Top K prediction accuracy)   + FFS: percentage based on predictions/measurements in a window (or whether or not correct according to one shot prediction/measurement   + FFS on other details including how to configure the resources for measurements * Option 2: The RSRP difference information between the measured L1-RSRP and predicted RSRP/L1-RSRP of predicted best beams   + FFS on statistical prediction performance metrics in a window or according to one shot   + FFS on the beam(s)/a beam set associated with the RSRP difference information     - E.g., RSRP difference information between predicted RSRP and measured L1-RSRP of the Top 1 or Top K predicted beam(s)     - E.g., RSRP difference information between the measured L1-RSRP of the Top-1 predicted beam and largest L1-RSRP over a beam Set (e.g. Set A or a set for monitoring).     - E.g. Difference of measured L1-RSRPs between genie-aided best beam index(es) from Set A (or a set for monitoring) and L1-RSRP of the predicted best beam index(es). * Option 3: The The probability/confidence of the predicted beam(s), ~~to be the Top 1 or Top K beam~~   + E.g. probability to be the Top 1 or Top K beam, e.g. confidence calculated from measured L1-RSRPs and predicted RSRPs for Set B * FFS: for all options, on other details including how to configure the resources for measurements,   + E.g., the associated CSI-ResourceConfigId of Set B and/or the time relationship to the predicted CSI can be indicated for the measurement of Set A. * For all options, study whether to report per sample (one-shot), report per set of samples (window) or statistical value. * Note: Whether to send the metric or an event based on the metric is a separate discussion   Other options are not precluded. |
| Fujitsu | We support Option 1 and Option 2. |
| New H3C | OK in general |
| TCL | Support Option 1 and Option 2 |
| Xiaomi | Support Option 1 and Option 2. We don’t think one shot is reasonable.  As for Option 3, does it mean the output of the model is the probability of the predicted beam(s) to be the Top 1 or Top K beam? and the performance metric is the probability? In this case, it is not necessary to perform measurement on Set A? |
| ZTE | For Option 2, the wording of the bullet cannot cover the second example in the sub sub-bullet，where the measured RSRPs between the predicted Top-1 beam and the genie-aided Top-1 beam are compared. To make it simple, the ending part can be deleted.  For Option 3, per our understanding, probability information only reflects the difference among multiple beam qualities, but may not be used for monitoring. Evaluations showing the relationship between the probability information and beam prediction accuracy are deficient. It may be possible that a beam prediction model output a very large probability for a incorrect Top-1 predicted beam, which would mislead the monitoring result determination. Besides, consider that even with exactly same model inputs, different beam prediction models can output very different probability values. Thus, it becomes difficult to determine a threshold value for monitoring purpose that can be applied to all UE-sided models.  Therefore, we suggest the following revisions for Option 2 and 3.   * Option 2: The RSRP difference information between the measured L1-RSRP and predicted RSRP   + FFS on statistical prediction performance metrics in a window or according to one shot   + FFS on the beam(s)/a beam set associated with the RSRP difference information     - E.g., RSRP difference information between predicted RSRP and measured L1-RSRP of the Top 1 or Top K predicted beam(s)     - E.g., RSRP difference information between the measured L1-RSRP of the Top-1 predicted beam and largest L1-RSRP over a beam Set (e.g. Set A or a set for monitoring). * Option 3: The The probability of the predicted beam(s) to be the Top 1 or Top K beam |
| Qualcomm | The second FFS on option 1 is applicable to all the options, hence moved to the end.  The following example mentioned under Option 2 is a totally different concept and cannot be categized under Option 2.   * + - E.g., RSRP difference information between the measured L1-RSRP of the Top-1 predicted beam and largest L1-RSRP over a beam Set (e.g. Set A or a set for monitoring).   Fundamentally, Option 2 is comparing a predicted RSRP to a measured RSRP, but the above example is comparing the *measured* RSRPs across two beams. Suggest updating as a separate option, as outlined below.  In our view, combination of metrics based on beam prediction accuracy and measured RSRP difference can be very beneficial. For the metrics based on beam prediction accuracy, we say the Top-1 predicted beam from Set A MUST be within Top-K predicted beams, whereas another option (Option 4 below) says as long as the RSRP difference between the Top-1 predicted beam from Set A (or best RSRP from Top-K predicted beams from Set A) is within a threshold compared to best RSRP from Set A, then this is considered a success event.  It is not clear how option 3 can be used for monitoring, suggest removing it unless clarification is made. It is very important to highlight the fact that the performance monitoring set in the options mentioned here span the entire Set A. Realistically, materializing such an assumption may be challenging in practice, and only a subset of beams from Set A may be configured for measurement, or UE may be able to measure only a subset of beams from Set A. The question is: how should the metrics be defined in such scenarios? Which is scenarios in which we do not know the ground truth measurements fully.  Updated Proposal 2.2  For BM-Case1 and BM-Case2 with a UE-side AI/ML model, for Option 2 (UE-assisted performance monitoring), further study the performance metric(s), as well as the corresponding report content(s) (e.g., the metric(s) itself or event(s) based on the metric(s)) and procedure(s), including:   * Option 1: The beam prediction accuracy related information (e.g., Top 1 or Top K prediction accuracy)   + FFS: percentage based on predictions/measurements in a window or whether or not correct according to ~~one shot~~ per-instance prediction/measurement   + FFS on other details including how to configure the resources for measurements * Option 2: The RSRP difference information between the measured L1-RSRP and predicted RSRP   + FFS on statistical prediction performance metrics in a window or according to one shot   + FFS on the beam(s)/a beam set associated with the RSRP difference information     - E.g., RSRP difference information between predicted RSRP and measured L1-RSRP of the Top 1 or Top K predicted beam(s)     - E.g., RSRP difference information between the measured L1-RSRP of the Top-1 predicted beam and largest L1-RSRP over a beam Set (e.g. Set A or a set for monitoring). * Option 3: The The probability of the predicted beam(s) to be the Top 1 or Top K beam * Option 3: The RSRP difference information between the measured L1-RSRP from Top-1 predicted beam (or highest measured RSRP from Top-K predicted beams) from Set A and measured RSRP of Top-1 measured beam ID from Set A. * Option 4: The beam prediction accuracy related information with an RSRP margin (e.g., Top 1 or Top K prediction accuracy)   + For each instance, if measured RSRP of Top-1 predicted beam (or highest measured RSRP from Top-K predicted beams) from Set A is within a margin compared to RSRP of Top-1 measured beam ID from Set A, beam prediction accuracy is fulfilled for that instance.   + FFS: percentage based on predictions/measurements in a window or whether or not correct according to per-instance prediction/measurement. * FFS on other details including how to configure the resources for measurements * Other options are not precluded. * Note: For all the above metrics it is assumed that the performance monitoring set is equal to Set A.   + FFS: how to define metrics when the performance monitoring set is a subset of Set A. |
| InterDigital | 1. difference between option 1 and option 3 is not clear to us. Prefer to delete Option 3.  2. “whether or not correct according to one shot prediction/measurement” is not clear. What is the intention?  3. For option 2, the RSRP difference should be derived between the same beam. If the best beam for prediction and the best beam for measure is different, we don’t think it’s valid information. For simplicity, prefer to delete the examples.  For BM-Case1 and BM-Case2 with a UE-side AI/ML model, for Option 2 (UE-assisted performance monitoring), further study the performance metric(s), as well as the corresponding report content(s) (e.g., the metric(s) itself or event(s) based on the metric(s)) and procedure(s), including:   * Option 1: The beam prediction accuracy related information (e.g., Top 1 or Top K prediction accuracy)   + FFS: percentage based on predictions/measurements in a window or ~~whether or not correct~~ according to one shot prediction/measurement   + FFS on other details including how to configure the resources for measurements * Option 2: The RSRP difference information between the measured L1-RSRP and predicted RSRP   + FFS on statistical prediction performance metrics in a window or according to one shot   + FFS on the beam(s)/a beam set associated with the RSRP difference information     - E.g., RSRP difference information between predicted RSRP and measured L1-RSRP of the Top 1 or Top K predicted beam(s)     - E.g., RSRP difference information between the measured L1-RSRP of the Top-1 predicted beam and largest L1-RSRP over a beam Set (e.g. Set A or a set for monitoring). * Option 3: The The probability of the predicted beam(s) to be the Top 1 or Top K beam * Other options are not precluded. |
| Ericsson | Support, we don’t see a need for the FFS, these are anyway covered in the main bullet.  The RSRP difference of the measured RSRP of the predicted Top-1 beam, and the measured RSRP of the actual Top-2 should for clarity be listed in a separate option.  Updated proposal  For BM-Case1 and BM-Case2 with a UE-side AI/ML model, for Option 2 (UE-assisted performance monitoring), further study the performance metric(s), as well as the corresponding report content(s) (e.g., the metric(s) itself or event(s) based on the metric(s)) and procedure(s), including   * Option 1: The beam prediction accuracy related information (e.g., Top 1 or Top K prediction accuracy) * Option 2: The RSRP difference information between the measured L1-RSRP and predicted RSRP of the Top 1 or any of the Top K predicted beam(s) * Option 3: RSRP difference information between the measured L1-RSRP of the Top-1 predicted beam and largest L1-RSRP over a beam Set (e.g. Set A or a set for monitoring). * Option 4: The The probability of the predicted beam(s) to be the Top 1 or Top K beam * Other options are not precluded. |
| LG | We also prefer to focus on Option 1 and Option 2 which have solid support. Regarding Option 3, we have similar view as DOCOMO that Option 1 covers Option 3 indeed. |
| CATT | The beam prediction accuracy related KPIs include Top-1 beam prediction accuracy, Top-K/1 beam prediction accuracy, Top-1/K beam prediction accuracy and Top-1 beam prediction accuracy within 1 dB margin. We think all these KPIs can be used for performance monitoring, and the specific metric can be configured by the network. Compared to Top1 or Top K beam prediction accuracy, Top-1/K beam prediction accuracy and Top-1 beam prediction accuracy within 1 dB margin can tolerance. Although it is not the Top-l beam, it can still meet the transmission requirements. Our suggestion is as following:   * Option 1: The beam prediction accuracy related information (e.g., Top 1 or Top K prediction accuracy, Top-1/K beam prediction accuracy,Top-1 beam prediction accuracy within 1 dB margin)   + FFS: percentage based on predictions/measurements in a window or whether or not correct according to one shot prediction/measurement   + FFS on other details including how to configure the resources for measurements   For Option 2, the L1-RSRP difference of beams in Set B can be also used to reflect the decrease of beam prediction accuracy. If L1-RSRP difference of beams in Set B is used as the performance metric, the gNB only needs to transmit RS with Set B beams for model inference and performance monitoring, and no additional RS transmission is needed.  But, some companies have concern on that some AI/ML model can’t use this method because their model can’t obtain predicted L1-RSRP of set B. We think the L1-RSRP difference of beams in Set B based performance monitoring can be as a UE capability. As a results, Our suggestion is as following:   * + FFS on the beam(s)/a beam set associated with the RSRP difference information     - E.g., RSRP difference information between predicted RSRP and measured L1-RSRP of the Top 1 or Top K predicted beam(s)     - E.g., RSRP difference information between the measured L1-RSRP of the Top-1 predicted beam and largest L1-RSRP over a beam Set (e.g. Set A or a set for monitoring).     - E.g., RSRP difference information between the predicted RSRP and the measured L1-RSRP of the set B beams |
| vivo | We are ok to discuss the details on Alt 1 and Alt 2.  For Alt 3, we think it is more relevant with the option of having probability as reporting content. We can discuss this after the decision of having probability in inference reporting content. |
| Nokia | We have done enough discussions on metrics from Rel-18, and let’s go ahead and agree on fundamental directions. Also, we do not fully understand the listed FFS items. Those can be discussed later.  Updated proposal  For BM-Case1 and BM-Case2 with a UE-side AI/ML model, for Option 2 (UE-assisted performance monitoring), ~~further study the~~ support following performance metric(s), ~~as well as the corresponding report content(s) (e.g., the metric(s) itself or event(s) based on the metric(s)) and procedure(s), including:~~   * Option 1: The beam prediction accuracy related information (e.g., Top 1 or Top K prediction accuracy)   + FFS: configurations/indications associated with calculating the metric, monitoring time durations, monitoring RS resources, and reporting format   + FFS: percentage based on predictions/measurements in a window or whether or not correct according to one shot prediction/measurement   + FFS on other details including how to configure the resources for measurements * Option 2 (applicable only for RSRSP prediction): The RSRP difference information between the measured L1-RSRP and predicted RSRP   + FFS: configurations/indications associated with calculating the metric, monitoring time durations, monitoring RS resources, and reporting format   + FFS on statistical prediction performance metrics in a window or according to one shot   + FFS on the beam(s)/a beam set associated with the RSRP difference information     - E.g., RSRP difference information between predicted RSRP and measured L1-RSRP of the Top 1 or Top K predicted beam(s)     - E.g., RSRP difference information between the measured L1-RSRP of the Top-1 predicted beam and largest L1-RSRP over a beam Set (e.g. Set A or a set for monitoring). * Option 3: The The probability of the predicted beam(s) to be the Top 1 or Top K beam * Other options are not precluded. |
| KDDI | We support Option 1 and Option 2. |
| Panasonic | We could not see any difference between Option 1 and Option 3 as mentioned similarly by other companies. We would like to hear proponents to elaborate about the difference. Otherwise, Option 3 can be removed. |
| Spreadtrum | We support Option 1 and Option 2. And we also think the performance metrics should be determined firstly. |
| CMCC | Suggest to prioritize option 1. |
| Kyocera | The key difference between options 1 and 3 is that option 3 relies on probability information derived from the output of the ML model. This probability information is based on the model's training and generalization, rather than a comparison to ground truth. For this reason, we support deprioritizing option 3. |
| Sony | We support options 1 and 2. For option 2, we believe it's unreasonable to consider only one shot, at least considering a time window. As for option 3, it’s not clear the probability information is derived from the model output or calculated by the UE. If it's based on the model output, it shouldn't be used as a metric to evaluate the model's performance. If the UE calculates it, how does this differ from option 1? |
| Sharp | We are generally fine with option 1 and option 2, which provides evaluation based on prediction and measurement. While Option 3 is more like the contents of inference report, which is still FFS in previous agreement. |
| Futurewei | We are open to further study Option 1. Suggest deprioritizing Option 2 and Option 3. |
| NEC | For Option 2 - First example, to our understanding, the main consideration is to reduce overhead during performance monitoring. But actually, performance monitoring should not a frequent behavior for a trained AI/ML model. Therefore, we prefer to perform measurements on the entire or part of Set A rather than only considering the predicted Top-1 or Top-K predicted beam(s).  For Option 2 - Second example, we suggest the following update:   * E.g., RSRP difference information between the measured L1-RSRP of the Top-1 predicted beam and ~~largest L1-RSRP over~~ the measured L1-RSRP of the Top 1 genie-aided beam of a beam Set (e.g. Set A or a set for monitoring)   Additionally, in our view, for regression-based model, the AI output should be L1-RSRPs of all beams of the Set A, so naturally, the following RSRP difference information should also be considered.  E.g., RSRP difference information between and predicted L1-RSRPs and measured L1-RSRPs of all beams in a beam set (e.g., Set A or a set for monitoring) |
| Lenovo | We are fine to focus on Option 1 and Option 2. As some companies pointed that Option 3 may be covered by Option 1 when the prediction accuracy related information is set as probability. Another comment is that does the probability is the model output of the corresponding beam? If the answer is yes, we can’t understanding why the performance monitoring can be supported by without any measurement on the beam set A. |

##### (FL1) Proposal 2.2a:

For BM-Case1 and BM-Case2 with a UE-side AI/ML model, for Option 2 (UE-assisted performance monitoring), further study the performance metric(s), as well as the corresponding report content(s) ~~(e.g., the metric(s) itself or event(s) based on the metric(s))~~ and procedure(s), including:

* Option 1: Top 1 or Top K beam prediction accuracy (with or without margin) by comparing the prediction results and the Top 1 or Top K beam based ~~on measurements of Set A~~ measurement resource set for monitoring
* Option 2a: The RSRP difference information between the measured L1-RSRP and predicted RSRP of the Top 1 or Top K predicted beam(s)
  + Note: this is can only applicable when a model can predict RSRP
* Option 2b: The RSRP difference information between the measured L1-RSRP of the Top-1 or Top K predicted beam and Top 1 or Top K largest L1-RSRP over a beam Set (e.g. Set A or a measurement resource set for monitoring).
* Option 3: The probability information of the predicted beam(s) to be the Top 1 or Top K beam
  + Note: this is can only applicable when a model can provide probability information as model output
* Option 4: The confidence calculated from measured L1-RSRPs and predicted RSRPs for Set B
  + Note: this is can only applicable when a model can calculate confidence information as model output
* Other options are not precluded.
* FFS: for option 1/2a/2b, on other details including how to configure the resources for measurements,
* For all options, study whether to report per sample (one-shot), report per set of samples (window) or statistical value.
* Note: Whether to send the metric or an event based on the metric is a separate discussion

|  |  |
| --- | --- |
| Companies | View |
| FL | I personally have many questions and concerns.  For option 1, how to define the metric without asking UE to always measure full Set A?  Similar for option 2b. but it may work if configuring a subset of Set A. But still, the subset of Set A is trustable enough, why need to predict it? Why not directly select beam based on subset of Set A?  Option 2a may work, which requires: A) NW configure UE to measure predicted Top 1/K beams. The RSRP difference somehow can provide information on whether the model works well or not. I think there are some contributions in SI phase.  Similarly, option 3, the probability information can also provide some information on whether AI trust itself about the prediction results. It has also been analyzed is SI phase from some contributions.  The issue for Option 2a and Option 3 can only be applicable for certain type of model.  For option 4, if this is based on “Set B”, it is even more restricted on UE implementation, i.e., how to label a model. I don’t think we should spend time to define a metric does not provide clear gain but with strong limitation.  I suggest proponent companies to explain and resolve the above concerns first. |
| TCL | Firstly, the term of beam prediction accuracy in Option 1 should be clarified. In my understanding, the metric of beam prediction accuracy could be generally the follows: (1) The difference in predicted best K beam indexes and the measured ones; (2) The difference in predicted and measured RSRPs of the best K beams, i.e., the Option 1 and Option 2, or the combination of the two. Option 1 and option 2 are the most trustworthy as they are derived from the ground truth values, i.e., the measured value of the measurement resource set. In contrast, option 3 is less reliable since the probabilities is also the predictive output of the model, it is not credible to use a predictive value, i.e., the probability, to determine the performance of another predictive value, i.e., the Top K beams. The Option 4 in proposal 2.2a has same issue.  However, I think 3GPP should not specify which type of AI/ML model to be used, as it depends on the vendor to implement. Among all options only Option 1 is not subject to AI/ML model type. Then I suggest in current stage we list all possible options and set Option 1 as the baseline or mandatory.  Updated proposal:  For BM-Case1 and BM-Case2 with a UE-side AI/ML model, for Option 2 (UE-assisted performance monitoring), further study the performance metric(s), as well as the corresponding report content(s) ~~(e.g., the metric(s) itself or event(s) based on the metric(s))~~ and procedure(s). The performance metrics are based on the Top 1 or K beam prediction accuracy, including:   * Option 1: Comparing the prediction results, e.g., the predicted beam indexes, and the Top 1 or Top K beam based ~~on measurements of Set A~~ measurement resource set for monitoring : * Option 2a: The RSRP difference information between the measured L1-RSRP and predicted RSRP of the Top 1 or Top K predicted beam(s)   + Note: this is can only applicable when a model can predict RSRP * Option 2b: The RSRP difference information between the measured L1-RSRP of the Top-1 or Top K predicted beam and Top 1 or Top K largest L1-RSRP over a beam Set (e.g. Set A or a measurement resource set for monitoring). * Option 3: The probability information of the predicted beam(s) to be the Top 1 or Top K beam   + Note: this is can only applicable when a model can provide probability information as model output * Option 4: The confidence calculated from measured L1-RSRPs and predicted RSRPs for Set B   + Note: this is can only applicable when a model can calculate confidence information as model output * Option 5: The difference between the predicted and measured strongest Top K beams, and the RSRP difference information between the predicted and measured L1-RSRP of the overlapped part of the predicted and measured Top K beams. * Other options are not precluded. * FFS: for option 1/2a/2b/5, on other details including how to configure the resources for measurements, * For all options, study whether to report per sample (one-shot), report per set of samples (window) or statistical value. * Note: Whether to send the metric or an event based on the metric is a separate discussion |
| Xiaomi | Support Option 1 and Option 2a.  For Option 1, we think the measurement resource set should be set A.  For Option 2b, if the resource ID is different, it may be meaningless to compare the RSRP difference.  For Option 3, we don’t think it can be used for monitoring without measurement on Set A. |
| CMCC | Support Option 1.  For Option 3, we think the probability means the predicted probability of one beam to be top1 beam. We confuse how can Option 3 be a monitoring metric.  For option 4, the confidence information is output of AI model, the definition of this confidence information is not clear to us. |
| Kyocera | For option 1, in the performance monitoring phase, the NW should transmit RSs beamformed by the spatial TX filters in Set A of beams on the “measurements resource set for monitoring”. Hence, as per our understanding for option 1, if Set A resources are configured for the UE, then the “measurements resource set for monitoring” should be the same as “resources for Set A of beams”. However, when Set A resources are not configured for the UE, then it makes sense to say “measurements resource set for monitoring” since they could be any set of resources dedicated for monitoring.  Our understanding for option 2 is as follows:   * For option 2-a, the UE reports the RSRP difference between the predicted RSRP and the L1-RSRP from the corresponding resource. For example, if the ML model predicts CRI “k” and its predicted RSRP, then the RSRP difference is taken with respect to the L1-RSRP measurement of resource “k” in the resource set configured for monitoring. In that case, the size of this resource set should be equal to the size of Set A beams. * For option 2-b, the UE reports the RSRP difference between the predicted RSRP and the maximum L1-RSRP of the configured resources for performance monitoring. For example, if the ML model predicts CRI and its predicted RSRP, then the RSRP difference is taken with respect to the maximum measured L1-RSRP in the resource set configured for monitoring, i.e., can be the L1-RSRP of resource . In that case, this resource set can be a subset of Set A. However, we do not see the motivation for such configuration, i.e., monitoring set is a subset of Set A. However, if the monitoring resource set is a subset of Set A, it means that the ML model outputs is always compared to a specific set of beams (good beams) that are part of all Set A beams (all beams). If we know some good beams from the beginning, it is unclear for us why the model should be trained on all beams. Hence, for option 2-b, we see no motivation to say that the set configured for monitoring is a subset of Set A. |

### Issue #2.3(onhold): For UE sided model, FFS whether to trigger the report based on event(s) for Option 1 and/or Option 2

Potential contents in RAN 1 #117:

* Event-1: Predicted beam accuracy of the set of predicted beams being below a threshold accuracy/The measured Top-1 or Top K beam(s) of Set A and the predicted Top-1 or Top K beam(s) of Set A are different
  + FFS on whether one shot or statistical results in a given window is used
  + Comments from FL: One shot is not reliable. I think in a window should be considered.
  + Supported by: Intel (Xdb margin), Nokia, Sharp, samsung
* Event-2: The measured L1-RSRP of one set of beams is lower than a threshold.
  + Comments from FL: isn’t this similar as BFD?
  + Supported by: Intel (Xdb margin)
* Event-3: The probability information of Top-1 or Top K beam of Set A is lower than a threshold, or difference comparing to previous is larger than a threshold
  + FFS on how to define the probability information
    - #1: The probability information of predicted Top 1
    - #2: The probability information of each or sum of predicted Top Top-K beams.
  + Comments from FL: this may be straightforward.
  + Supported by: OPPO,Samsung
* Event-4: The L1-RSRP difference between the measured Top-1 or Top K beam(s) of Set A and predicted Top-1 or Top K beam(s) of Set A are larger than a threshold value
  + FFS on RSRP difference information: e.g., RSRP difference, whether RSRP difference is higher than a threshold, all or part of RSRP difference
  + FFS on whether/how define the associated beams for RSRP difference information, e.g.,
    - #1: of a set of beams configured by NW
      * FFS on whether/how to handle the case if the configured beams are not the predicted Top 1 or Top K
    - #2: of predicted Top 1 or Top K beams
      * UE is not required to report the RSRP difference information if the configured beam is not predicted Top 1 or Top K beams
    - #3: RSRP difference between predicted Top 1 or Top K beams, and Top 1 or Top K beams of a set of beams (e.g., full or subset of Set A, same or different as predicted Top 1 or Top K beams)
  + FFS on how to configure resources to obtain the measured L1-RSRP
  + Comments from FL: Similar as Option C for type 1 performance monitoring
  + Supported by: Intel ,Samsung
* Event-5: consecutive N1 times of KPI/BLER less than threshold, or accumulated N1 times of KPI/BLER less than threshold within a time period.
  + Supported by: CMCC, Nokia

Event-6: Predicted L1-RSRP of the set of predicted beams being below a threshold value.

* + Supported by: Nokia
* Event-7: The Counter is incremented if the predicted Top K beam results differ from the actual measured results and/or if the L1-RSRP difference exceeds a predefined value
  + Supported by: ETRI

# Configuration of RS for Set A and Set B for NW sided model(onhold)

|  |
| --- |
| Agreement  For network-sided AI/ML model for BM-Case1 and BM-Case2,   * support using existing CSI framework for configuration of Set A as the starting point * support using existing CSI framework for configuration of Set B as the starting point   Note: Purpose, such as above “For NW-sided model, for BM-Case1 and BM-Case2” and “Set A” and “Set B”, will not be specified in RAN 1 specifications |

### Summary from the contributions

|  |  |  |
| --- | --- | --- |
| Companies | | Proposals |
| CMCC [5] | | Proposal 4: Regarding data collection for NW-side model, following options can be considered for the configuration of Set A:   * Option 1: The UE capability of the maximum number of the RS per resource set for the RSRP measurement can be enhanced. * Option 2: Multiple resource sets can be supported in one CSI-ResourceConfig   Proposal 5: Regarding training data collection for NW-side model, the configuration of multiple set B or multiple Set B patterns can be supported.  Proposal 15: For NW-sided model, for BM-Case 2 without sliding measurement window, CSI report framework needs enhancement to configure the resource set for measurement.  Proposal 16: For NW-sided model, for BM-Case 2 with sliding measurement window, measurement window is determined by gNB. |
| Intel [6] | | For a network-side AI/ML model, for BM-Case 1/2, implicit configuration of set B and/or set A (if required) for AI/ML model input can be used based on existing CSI-ResourceConfig and CSI-ReportConfig frameworks. |
| ZTE [7] | | Proposal 5: 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 6: For the collection of both model input and model label data at the NW side, one RS resource set can be configured to the UE with indication of necessary assistance information, such as mapping of Set A and Set B. |
| Vivo [9] | | Proposal 22: For inference, for NW-side model, support gNB to configure a resource list including multiple Set B patterns and active resources of one Set B pattern from the list by signaling during the inference phase.  Proposal 23: For NW-side model, there is no motivation to configure Set B and Set A simultaneously in one report.  Proposal 24: For data collection and inference, for NW-side model, support to use quasi-best Rx beam for Set A measurement, where quasi-best Rx beam is derived from P3 measurement on a small number of Tx beams from Set A. |
| Fujitsu [11] | Proposal 5:   * For training data collection, the reference signals same as Set B should be configured to obtain the model input data. And the reference signals same as Set A should be configured to obtain the ground truth data. * If Set B is subset of Set A, then only the reference signals of Set A are configured. * If Set B is different from Set A, then the reference signals of both Set A and Set B should be configured to the UE.   Proposal 6:   * RAN1 to further discuss whether the same or separate reference signal configuration among training data collection, model inference and performance monitoring could be applied.   Proposal 7:   * Regarding training data collection, RAN1 to further discuss the reference signal configuration for different sub-use cases. | |
| Nokia [19] | Proposal 22: For BM-Case1 and BM-Case2 with the NW-sided model, to enable the NW-sided performance monitoring, further discuss following variants,   * Case1: No enhancement is needed to support NW-sided performance monitoring.   + E.g., the NW can use a different CSI report to get beam measurements/reporting for a monitoring RS resource set (as NW prefer) within the legacy CSI reporting framework. * Case 2: NW is using the same CSI reporting configuration for monitoring and inference.   + Monitoring RS resource set is configured/indicated separately from Set B. For the monitoring RS resource set, the NW may configure separate reporting timelines and reporting quantities. | |
| ETRI [23] | Proposal 9: For BM-Case2, support the configuration of multiple Resource Sets for Set A. | |
| Xiaomi [34] | Proposal 3-1: For data collection of NW-side AI/ML model training, support to define a time window or a data size for each report with more than one data sample.  Proposal 3-5: Both two separate CSI-ReportConfigs and one CSI-ReportConfig can be supported for set B and set A configuration for data collection for NW-side AI/ML model training.  Proposal 3-6: If one CSI-ReportConfig is used for set B and set A configuration for data collection for NW-side AI/ML model training, consider to support more than one reportquantity in one CSI-ReportConfig. | |
| LGE [36] | Proposal #3: Consider extending sub-configuration based Rel-18 NES mechanism for Set B beam measurement and reporting.   * Different Set A and/or Set B may be associated with each sub-configuration * Different Set B patterns for a specific Set B may be associated with each sub-configuration | |
| MTL [39] | Proposal 19: For NW-side model inference BM-Case2, for periodic Set B RS resources, study the following alternatives for configuring the observation window length to UE:   * Note: the observation window length is how many Set B measurements need to be reported * Alt1: observation window length is configured within LCM framework   + the length of observation window should be one of the conditions to BM functionalities * Alt2: observation window length is configured within CSI-framework   + Alt A: explicitly configured in csi-ReportConfig   + Alt B: implicitly configured in csi-ReportConfig     - P/SP report: indicated through ReportPeriodicityAndOffset of the report and ResourcePeriodicityAndOffset of the associated RS resources of Set B     - AP report: indicated by the triggered slot and report slot   Proposal 20: For NW-side model inference BM-Case2, for aperiodic Set B RS resources and aperiodic report triggering, confirm that there is need to further configure observation and prediction window information to UE  Proposal 21: For NW-side model inference, support reporting multiple time instances of Set B measurements within one report. Further study on whether/how to explicitly and/or implicitly include corresponding time information in the report.  Proposal 22: For NW-side model inference, same design of report and resource configurations can be used for BM Case1 and BM Case2 | |

### Issue #3.1: Whether to configure multiple resource sets associated to one L1 beam report for NW sided model

Whether to configure multiple resource sets associated to one L1 beam report for NW sided model,

* for Set A and Set B to enable, one report beam ID information to one report and L1-RSRP for another report
* for BM-Case 2? If yes, please explain how to handle time stamp information in a report conf.

### Issue #3.2 Others

Whether using current configuration for Set A measurement and report have any restriction to gNB scheduling and/or require additional capability from UE? considering combination of AP, SP, P of resource config and report config. If, yes, any enhancement to resolve it?

For BM case 2, using current configuration for Set B and Set A (for data collection and or training) measurement and report have any restriction to gNB scheduling and/or require additional capability from UE?

Please provide your detailed analysis to the above questions/issues, observation, and proposals.

CMCC:

Proposal 15: For NW-sided model, for BM-Case 2 without sliding measurement window, CSI report framework needs enhancement to configure the resource set for measurement.

Proposal 16: For NW-sided model, for BM-Case 2 with sliding measurement window, measurement window is determined by gNB.

# Configuration of RS for Set A and Set B for UE-sided model

|  |
| --- |
| Review of current NR CSI framework:  CSI-MeasConfig ->CSI-ReportConfig ->resourcesForChannelMeasurement (and resources for other purposes) => – CSI-ResourceConfig  ->Reporting related configuration  The IE CSI-ResourceConfig defines a group of one or more NZP-CSI-RS-ResourceSet, CSI-IM-ResourceSet and/or CSI-SSB-ResourceSet.  csi-RS-ResourceSetList CHOICE {  nzp-CSI-RS-SSB SEQUENCE {  nzp-CSI-RS-ResourceSetList SEQUENCE (SIZE (1..maxNrofNZP-CSI-RS-ResourceSetsPerConfig)) OF NZP-CSI-RS-ResourceSetId  OPTIONAL, -- Need R  csi-SSB-ResourceSetList SEQUENCE (SIZE (1..maxNrofCSI-SSB-ResourceSetsPerConfig)) OF CSI-SSB-ResourceSetId OPTIONAL -- Need R  },  csi-IM-ResourceSetList SEQUENCE (SIZE (1..maxNrofCSI-IM-ResourceSetsPerConfig)) OF CSI-IM-ResourceSetId  },  The IE *NZP-CSI-RS-ResourceSet* is a set of Non-Zero-Power (NZP) CSI-RS resources (their IDs) and set-specific parameters.  The IE *NZP-CSI-RS-Resource* is used to configure Non-Zero-Power (NZP) CSI-RS transmitted in the cell where the IE is included, which the UE may be configured to measure on (see TS 38.214 [19], clause 5.2.2.3.1). A change of configuration between periodic, semi-persistent or aperiodic for an *NZP-CSI-RS-Resource* is not supported without a release and add.  The IE *CSI-RS-ResourceMapping* is used to configure the resource element mapping of a CSI-RS resource in time- and frequency domain.  maxNrofCSI-ReportConfigurations INTEGER ::= 48 – Maximum number of report configurations  maxNrofCSI-ReportConfigurations-1 INTEGER ::= 47 – Maximum number of report configurations minus 1  maxNrofCSI-ResourceConfigurations INTEGER ::= 112 – Maximum number of resource configurations  maxNrofCSI-ResourceConfigurations-1 INTEGER ::= 111 – Maximum number of resource configurations minus 1  maxNrofNZP-CSI-RS-Resources INTEGER ::= 192 – Maximum number of Non-Zero-Power (NZP) CSI-RS resources  maxNrofNZP-CSI-RS-Resources-1 INTEGER ::= 191 – Maximum number of Non-Zero-Power (NZP) CSI-RS resources minus 1  maxNrofNZP-CSI-RS-ResourcesPerSet INTEGER ::= 64 – Maximum number of NZP CSI-RS resources per resource set  maxNrofNZP-CSI-RS-ResourceSets INTEGER ::= 64 – Maximum number of NZP CSI-RS resource sets per cell  maxNrofNZP-CSI-RS-ResourceSets-1 INTEGER ::= 63 – Maximum number of NZP CSI-RS resource sets per cell minus 1  maxNrofNZP-CSI-RS-ResourceSetsPerConfig INTEGER ::= 16 – Maximum number of resource sets per resource configuration  maxNrofNZP-CSI-RS-ResourcesPerConfig INTEGER ::= 128 – Maximum number of resources per resource configuration  BRF configuration  BeamFailureRecoveryConfig-> candidateBeamRSList-> PRACH-ResourceDedicatedBFR-> BFR-SSB-Resource/ BFR-CSIRS-Resource(->NZP-CSI-RS-ResourceId) |
| Conclusion  For UE sided model at least for inference, for measurement, the configuration of Set B,   * take the current CSI framework as the starting point   Agreement  For UE-sided model at least for BM Case-1, *CSI-ReportConfig* is used for the configuration of inference results reporting   * FFS on the details in the *CSI-ReportConfig*, at least considering:   + Alt 1: one *CSI-ResourceConfigId* is configured for Set B     - FFS: how UE can determine the information about set A   + Alt 2: one *CSI-ResourceConfigId* is configured for both Set A and Set B     - FFS: How to configure resource set(s) for Set A and Set B in *CSI-ResourceConfig*   + Alt 3: two *CSI-ResourceConfigId* s are configured for Set A and Set B separately   + Alt 4: one *CSI-ResourceConfigId* is configured for Set B, Set A is configured using separate resource set(s) other than that represented by *CSI-ResourceConfigId*     - FFS: how to configure/indicate separate resource set(s) for Set A   + Note: separate *CSI-ReportConfig* for Set A and Set B are not precluded.   + Note: Not perform measurement for Set A and only perform measurement for Set B subject to the *CSI-ReportConfig*   + FFS on the association between Set A and Set B with or without additional IE * Other necessary configuration are not precluded. |

### Summary from the contributions

|  |  |
| --- | --- |
| Companies | Proposals |
| Futurewei [1] | Proposal 3: For Rel-19 AI/ML-based BM, for UE-sided model at least for BM Case-1, CSI-ReportConfig is used for the configuration of inference results reporting. On the details in the CSI-ReportConfig, further consider Alt 2 and Alt 3:   * Alt 2: one CSI-ResourceConfigId is configured for both Set A and Set B. * Alt 3: two CSI-ResourceConfigId s are configured for Set A and Set B separately.   Proposal 6: For Rel-19 AI/ML-based BM, for data collection for UE-sided AI/ML model of BM-Case1 and BM-Case2, support that NW provides/signals multiple possible configurations of DL RS transmission to the UE and the UE reports its supported/preferred one(s) out of the multiple configurations.  Proposal 7: For Rel-19 AI/ML-based BM, for data collection for AI/ML model of BM-Case1 and BM-Case2, support using RS ID as implicit indication of beam ID and reusing L1-RSRP reporting as much as possible. |
| Spreadtrum [2] | Proposal 1：For UE-side model, support UE to request the data collection and report training-related information, such as expected measurement resources, etc.  Proposal 2: For the configuration of inference results reporting for UE-side model, support Alt 2. |
| GOOGLE [3] | Proposal 10: AI/ML based beam prediction should not mandate the configuration of set A DL RS.   * For model inference, Alt1 should be supported where UE can derive the set A beam information based on the associated ID * Study overhead reduction mechanism for set A RS configuration   Proposal 11: For Set A RS, support the NW configures 1-port CSI-RS and UE shall not expect the NW multiplex any other DL RS in the same symbol with the set A RS.   * The NW transmits the set A RS based on IFDMA structure for fast UE beam refinement   Proposal 12: Support to configure aperiodic CSI-RS as set A RS, which is triggered by a group-cast DCI. |
| CMCC [5] | Proposal 7: Regarding data collection for UE-side model for training, UE can request for preferred set B.  Proposal 8: Regarding data collection for UE-side model for training, set B as AI model input is determined by gNB.  Proposal 9: Regarding data collection for UE-side model, the configuration method of set A for NW-side model can be reused.   * Option 1: The UE capability of the maximum number of the RS per resource set for the RSRP measurement can be enhanced. * Option 2: Multiple resource sets can be supported in one CSI-ResourceConfig   Proposal 10: For UE-sided model at least for BM Case-1, following two options are supported for the configuration of training data collection.   * + Option 1: one CSI-ResourceConfigId is configured for both Set A and Set B     - separate resource sets are configured for Set A and Set B in CSI-ResourceConfig   + Option 2: two CSI-ResourceConfigId s are configured for Set A and Set B separately   Proposal 11: Regarding to option 1 of training data collection for UE-sided model, indication of association of Set A and Set B can be based on RS ID or bitmap.  Proposal 12: Regarding to option 2 of training data collection for UE-sided model, indication of association of Set A and Set B can be based on RS ID.  Proposal 21: For UE-sided model at least for BM Case-1, CSI-ReportConfig is used for the configuration of inference results reporting   * FFS on the details in the CSI-ReportConfig, at least considering:   + Option 2: one CSI-ResourceConfigId is configured for both Set A and Set B     - separate resource sets are configured for Set A and Set B in CSI-ResourceConfig     - Set A may be a virtual set including RS ID and resource configuration, RS in Set A is not transmitted unless triggered by gNB     - indication of association of Set A and Set B can be based on RS ID or bitmap   + Option 3: two CSI-ResourceConfigId s are configured for Set A and Set B separately     - Set A may be a virtual set including RS ID and resource configuration, RS in Set A is not transmitted unless triggered by gNB     - indication of association of Set A and Set B can be based on RS ID   Proposal 22: For UE-sided model, for BM-Case 2 without sliding measurement window, CSI report framework needs enhancement to configure the resource set for measurement.  Proposal 23: For UE-sided model, for BM-Case 2 with sliding measurement window, measurement window is determined by UE. |
| Intel [6] | For a UE-side AI/ML model, for BM-Case 1/2, explicit configuration of set A and set B should be supported and the configuration may be based on UE capability and any UE-side conditions related to supported model and input/output types.  For a UE-side AI/ML model, for BM-Case 1/2, support Alt 4: One CSI-ResourceConfigId is configured for Set B, Set A is configured using separate resource set(s) other than that represented by CSI-ResourceConfigId.   * Set A is configured using a new IE which lists the resources which form the QCL sources for DL Tx beams which are mapped to the output of the UE-side model, e.g.,   + Alt. A: Set A is defined as a TCI-State list where each TCI state corresponds to a DL Tx beam and the qcl-info in the TCI state contains a source RS which provides the QCL source (Type D). This source RS can be an NZP-CSI-RS-ResourceId or SSB-Index.   + Alt. B: Set A is defined as a set of NZP-CSI-RS-ResourceId or SSB-Index and the UE can derive the TCI states based on the TCI state list configuration where the resources in set A can be QCL sources. * Set B may be configured via an association to set A or independently.   For a UE-side AI/ML model, for BM-Case 2, UE may be configured with an observation window by the network. The prediction window configuration may be based on UE capability on the length of the window that the UE-side model can support.  Proposal 14: For data collection for a UE-side AI/ML model, consider UE triggering for data collection from the network based on a configured set A of beams. |
| ZTE[7] | 1. For UE-sided model at least for BM Case-1, CSI-ReportConfig is used for the configuration of inference results reporting, and the following two alternatives for CSI-ReportConfig can be considered for potential down selection.  * Alt 1: one CSI-ResourceConfigId is configured for Set B * Alt 3: two CSI-ResourceConfigId s are configured for Set A and Set B separately   Proposal 19: If Set A and Set B are different, resources for Set A and resources for Set B are configured as separate resource sets, and the association between Set A and Set B can be established based on the same CSI report setting.  Proposal 20: If Set B is a subset of Set A, only resources for Set A is configured, and resources for Set B is indicated as a subset of Set A based on assistance information provided by the NW, such as the mapping between Set A and Set B in the form of bitmap.  Proposal 24: Support flexibly indication/activation/deactivation of arbitrary beams or beam subsets among all beams in Set A to reduce the signaling overhead for Set B configuration. |
| Ericsson [8] | Figure 1: How TCI state IDs are configured in NR for aPeriodic and periodic measurements   1. Proposal 4: For UE-sided models, regarding Set B configuration for UE inference, conclude that the NW does not need to indicate “Set B”, no need to further discuss alternative 1-4 for set B.    * UE can use any of the NW configured measurement resources as input to its model. 2. Proposal 5: For UE-sided models, regarding set A configuration for UE inference, further study the following method for how to configure/report set A,    * Set A is configured in a ResourceSet, containing the beam IDs of set A,    * Set A ResourceSet does not have any physical resources (e.g. NZP-CSI-Resource)    * Set A ResourceSet is activated for reporting using existing mechanisms for aPeriodic, periodic and semi-persistent scheduling    * The reporting mechanisms (e.g. UCI) and format (e.g. CRI, L1-RSRP,…) are reused 3. Proposal 6: For UE-sided model inference, enable NW to specify set A beam subset restriction similar to codebook subset restriction (CBSR) that is specified for CSI feedback 4. Proposal 8 For UE-side AI/ML model inference, for BM-Case2, further study how NW can configure UE to report inference results of N future time slots, including    * NW- selected time instances: NW configure UE with N time instances relative to the reference time    * UE-selected time instances: NW configure UE to select and report N time instances relative to the reference time, with corresponding predictions    * Reference time definition: For example, MAC-CE/DCI activation slot, or UE reporting slot    * Maximum future time instance that can be predicted (e.g. 80ms/160ms/320ms/640ms/800ms/others) |
| Vivo [9] | Proposal 13: For inference, for UE-side model, support UE to recommend some preferred Set B patterns which were trained during the UE-side model training phase.  Proposal 14: For inference, for UE-side model, support gNB to configure a resource list including multiple Set B patterns and activate resources of one Set B pattern from the list by signaling during the inference phase.  Proposal 15: For UE-side model, support to configure full and/or subset of Set A associated with predicted beam report to address interference issue.  Proposal 16: For inference, for UE-side model, Set A or resources in Set A should be indicated as virtual Set or virtual resource, which does not require measurement on these resources for the report.  Proposal 17: For inference, for UE-side model, support below potential options of different alternatives for the configuration of inference results reporting,   * Alt1-Opt1: one CSI-ReportConfig, only includes Set B, Set A without any configuration can be assumed by associated ID. * Alt1-Opt2: one CSI-ReportConfig, only includes Set B, Set A is configured out of the CSI-ReportConfig and associated to Set B * Alt1-Opt3: two CSI-ReportConfig, one CSI-ReportConfig includes Set B and another CSI-ReportConfig includes Set A. * Alt2: one CSI-ReportConfig and one CSI-ResourceConfig, the CSI-ResourceConfig includes Set B and Set A * Alt3-Opt1: one CSI-ReportConfig with two CMR, one CMR includes Set B and another CMR includes Set A * Alt3-Opt2: one CSI-ReportConfig with one CMR, the CMR includes Set B and Set A   Proposal 18: For inference, for UE-side model, suggest to postpone any consideration of Alt4 until the discussion on the configuration of Set A is fully completed.  Proposal 19: For inference, suggest to configure different beams of Set B for measurement corresponding to different historical periods for BM-Case 2.  Proposal 20: For inference, further study how to avoid unnecessary measurements and RS overhead for CSI-RS occasions existed in the prediction window for BM-Case 2. |
| OPPO[10] | Proposal 17: For UE-side model inference, configure Set A which may contain SSB resources and/or CSI-RS resource that UE doesn’t have to measure.  Proposal 18: For UE-side model inference, configure more than one Set Bs and Set As and then select/activate one pair of associated Set B and Set A.  Proposal 19: For UE-side model, two separate CSI-ResourceConfigIds can be configured within a CSI-ReportConfig (Alt 3).  Proposal 20: For UE-side model training/inference/monitoring of BM-Case2, specify (if necessary) the measurement window and prediction window in time domain. |
| Fujitsu [11] | Proposal 10:   * For inference operation of BM Case-1 with UE side model, regarding the configuration of Set A and Set B, support the following options: * Option 1: Set A and Set B are configured via different resource set which are contained in the same CSI resource setting. * Option 2: Set A and Set B are configured via different CSI resource setting (CSI-ResourceConfig) * Option 3: Set A and Set B are configured via different reporting setting (CSI-ReportConfig)   Proposal 22:   * For BM Case-2 with UE side model, regarding configuration of Set A and Set B, similar design framework as BM Case-1 with UE side model could be considered.   Proposal 23:   * For BM Case-2 with UE side model, RAN1 to discuss that the UE can report the preferred pattern for measurement and prediction, including the number of measurement instances, the number of prediction instances, the measurement interval, and the prediction interval.   Proposal 24:   * For BM Case-2 with UE side model, RAN1 to consider overhead reduction for the inference results reporting, e.g., differential L1-RSRP across multiple time instances. |
| CATT [12] | Proposal 1：For resource configuration of a large number of beams (e.g., Set A), study how to configure the resources when the number of beams is larger than the max number of CSI-RS resources can be configured to UE for L1-RSRP measurement.  Proposal 2：For resource configuration of a large number of beams (e.g., Set A), study how to trigger the aperiodic RS transmitted in several slots when the number of beams is larger than the number of resources can be configured to measure L1-RSRP within a slot.  Proposal 3: For BM-Case2, consider the following configuration enhancement:   * Multiple RS sets, each RS set corresponding to one measurement time instance, can be associated to the CSI report of reporting the inference results; * UE shall derive the predicted L1-RSRP value reported in uplink slot n based on the measurements of only the most recent, no later than the CSI reference resource, transmission occasion of each resource set.   Proposal 9: For inference of UE-sided model, the following alternatives can be considered for the CSI-ReportConfig used for the configuration of inference results reporting:   * Alt 1: one CSI-ResourceConfigId is configured for Set B, and Set A can be determined from the associated functionality/association id; * Alt 2: one CSI-ResourceConfigId is configured for both Set A and Set B; * Alt 3: two CSI-ResourceConfigId s are configured for Set A and Set B separately; * Alt 4: one CSI-ResourceConfigId is configured for Set B, and Set A is indicated by CSI-ReportConfigId. |
| Lenovo [14] | Proposal 3: For the configuration of Set A and Set B in a CSI report with beam prediction, considering the following options:   * Option 1: Each Set B is associated with a Set A beams and the Set B beams is configured by a CSI-ResourceConfigId associated with a CSI-ReportConfig * Option 2: Set A beams and Set B beams are explicitly configured by two resource set in one CSI-ResourceConfigId associated with a CSI-ReportConfig   Proposal 10: RS configuration should be enhanced to support discontinuous RS transmission and adaptation for BM-Case 2. |
| Interdigital. [16] | Proposal 1: Support Alt 2: one CSI-ResourceConfigId is configured for both Set A and Set B.  Proposal 2: Support one of the following alternatives for Set A and Set B configuration.   * Alt 1: RS resources for Set A including Set B are configured in a RS resource set. * Alt 2: RS resources for Set A is configured in one RS resource set and RS resources for Set B is configured in another RS resource set.   Proposal 3: For Option 1, if supported, support a mechanism indicating a type of RS resource (i.e., Set A, Set B or neither).   * Option 1-1: RRC configuration to indicate a type of each RS resource. * Option 1-2: Dynamic activation of a type of each RS resource.   Proposal 16: Indicating configuration information associated with Set B to UE should be supported.  Proposal 17: Support reporting of UE selected Set B based on a rule (e.g., subset of best measured beams). |
| Panasonic [17] | Proposal 4: For UE-sided model inference, support two CSI-ResourceConfigId s are configured for Set A and Set B respectively.  Proposal 5: For UE-sided model inference, support that a measurement window can be configured with the measurement resource set. |
| Nokia [19] | Proposal 3: Consider the following when configuring Set A beams in a beam prediction related CSI report,   * For BM-Case1 and BM-Case2, considering the case of Set B is a subset of Set A, support Alt2 (one CSI-ResourceConfigId is configured for both Set A and Set B) as a basic case to configure/indicate a second RS resource set associated with the CSI report configuration to consider as Set A. * For BM-Case1 and BM-Case2, considering the case of Set A and Set B are different, support the following options.   + Option 1: Configure/Indicate a second RS resource set associated with the CSI report configuration (Alt2: one CSI-ResourceConfigId is configured for both Set A and Set B).   + Option 2: UE determines Set A based on QCL relations between Set B and configured CSI-RS resources (Alt1: one CSI-ResourceConfigId is configured for Set).     - Option 2 may be applied when there is no second resource set configured by the NW * For BM-Case2, considering Set A and Set B are the same, the legacy RS resource set (resourcesForChannelMeasurement) applicable to both Set B and Set A. * For BM-Case1 and BM-Case2, consider Alt.4 as an optional alternative depending on the discussions related to data collection and performance monitoring. |
| Samsung [21] | Proposal 8. For UE-sided model at least for BM Case-1, for the association between Set A and Set B, introduce DL Tx IDs for the identification of downlink spatial domain transmission filter.   * Each beam in Set A is associated with an DL Tx ID * Each beam in Set B is associated with an DL Tx ID * Note: UE assumes the beams corresponding to the same DL Tx ID shares the same downlink spatial domain transmission filter. * FFS: the relationship between DL Tx ID and associated ID. |
| Transsion [22] | Proposal 3: Regarding the trigger/initiating data collection for UE-side model, support Option 2:   * Option 2: request from UE for data collection.   Proposal 5: Regarding the configuration of inference results reporting for UE-side model, support Alt2 and Alt3:   * Alt 2: one CSI-ResourceConfigId is configured for both Set A and Set B   + FFS: How to configure resource set(s) for Set A and Set B in CSI-ResourceConfig * Alt 3: two CSI-ResourceConfigId s are configured for Set A and Set B separately.   Proposal 6: Regarding configuration for the measurements for BM-Case 2, study to trigger aperiodic resource set in multiple time instances. |
| DOCOMO [25] | Proposal 2: Resources of Set A and Set B should be configured in CSI-ResourceConfig   * Resource of Set A should be configured for UE measurements aimed to training and performance monitoring. * Resources of Set B should be configured for UE measurements aimed to inference and performance monitoring.   Proposal 6: Support the following configuration under CSI-ReportConfig (Alt3 or Alt4)   * One CSI-ResourceConfigId is configured for Set B * Another CSI-ResourceConfigId or resource set ID (nzp-CSI-ResourceSetId/csi-SSB-ResourceSetId) is configured for Set A   Proposal 10: Following approaches can be applied to obtain measurements at multiple time instances for BM-Case2  ・Periodic and semi-persistent consecutive CSI-RS transmission occasions are used as measurements at multiple time instances  ・Number of transmission occasions per aperiodic CSI-RS resource is configured to enable measurements at multiple time instances |
| Sharp [26] | Proposal 5 For UE-side model at least for BM Case-1, CSI-ReportConfig is used for the configuration of inference results reporting and the following alternatives should be considered:   Alt.2: one CSI-ResourceConfigId is configured for both Set A and Set B   Alt.3: two CSI-ResourceConfigIds are configured for Set A and Set B, respectively   Alt 4: one CSI-ResourceConfigId is configured for Set B, Set A is configured using separate resource set(s) other than that represented by CSI-ResourceConfigId  Proposal 11 For BM Case 2, consider two alternatives to trigger an aperiodic inference report:   Alt.1: multiple NZP-CSI-RS resource sets/CSI-SSB-ResourceSets are associated with one CSI report configuration   Alt.2: apply multiple triggering offsets to a single aperiodic NZP-CSI-RS resource set |
| Qualcomm [27] | Proposal 6 For UE-sided beam prediction, at least for BM-Case 1, regarding the FFS on the details in the CSI-ReportConfig, support the following:   * Alt 2: one CSI-ResourceConfigId is configured for both Set A and Set B  Proposal 9 For beam prediction for UE-side AI/ML models, specify signalling details associated with transmission of reference signals for performance monitoring (that span entire Set A or subset of beams from Set A), helping UE to assess the performance of UE-side AI/ML models.   * FFS: Type of RS for performance monitoring purpose (periodic/semi-persistent/aperiodic) * FFS: details of the case in which RS for performance monitoring is a subset of Set A (e.g., how to determine the subset, its variability over time, and the signalling details for indicating the corresponding subsets) |
| Fraunhofer HHI, Fraunhofer IIS [29] | Proposal 10: Support Alt 2, one CSI-ResourceConfigId is configured for both Set A and Set B, and the Set B of beams is provided as a subset of the CSI resource configuration.  Proposal 15: For BM-Case2, for UE-sided models, introduce a configuration that allows the exclusion of certain past measurements for the inference. |
| ITL [31] | Proposal 6: It can be considered the reporting the preferred DL RS configurations for the data collection for UE side training when requesting training via UE signaling.  Proposal 15: For data collection for UE side model inference, consider UE to send a request for preferred DL RS configuration and/or DL RS transmission  Proposal 16: For UE side model inference, existing specifications should be the baseline for the configuration or triggering CSI-RS/SSB of Set B  Proposal 17: For UE side model inference, it is proposed to support that both resources for Set A and resources for Set B are configured as two separate resources   * FFS on additional signaling to indicate the association |
| Huawei/HiSi[33] | Proposal 8: For measurements configured to facilitate the AI/ML operations of a UE-side model, the purpose or implied UE behavior of the measurement configuration (e.g. training, inference, monitoring, non-AI/ML) needs to be indicated to the UE since the corresponding UE behavior or the content of the UE report may be different. E.g.,   * For purpose of training, implied behavior - no report needed. * For purpose of inference, implied behaviors - report the predicted beams/RSRPs. * For purpose of monitoring, implied behaviors - report the label, predicted output or a metric. * For purpose of non-AI/ML, implied behaviors - report the measured beams/RSRPs.   Proposal 9: For UE-side model at least for BM-Case 1, for the configuration of inference results reporting, at least consider Alt 3: two CSI-ResourceConfigIds are configured for Set A and Set B separately.   * The associated CSI-ResourceConfigId of Set A can be indicated in the CSI-reportConfig of Set B.   Proposal 10: For the data collection for the UE-side model under BM-Case 1/BM-Case 2, the mapping between Set B and Set A can be supported for the case when Set B is a subset of Set A, e.g. with a bitmap or a list of CRIs of Set B. |
| Xiaomi [34] | Proposal 3-7: Both explicit and implicit association between set B and set A can be supported for data collection for UE-side AI/ML model training.   * Implicit association:   + With same resourceConfig ID or reportConfig ID, resource set with lower set ID(s) is(are) for set B, the last one is for set A.   + With same resource set ID, resources with lower IDs are for set B and others are for set A. * Explicit association:   + Introduce resourceforModelInput and resourceforModelOutput in reportConfig.   + Introduce data set/association ID for each resource set.   + With same resource set ID, introduce indication to indicate which resources are for set B.   Proposal 4-3: For CSI-ReportConfig for the configuration of inference results reporting for UE-sided model at least for BM Case-1, support Alt 1 and Alt 4.   * Alt 1: one CSI-ResourceConfigId is configured for Set B   + FFS: how UE can determine the information about set A * Alt 4: one CSI-ResourceConfigId is configured for Set B, Set A is configured using separate resource set(s) other than that represented by CSI-ResourceConfigId   + FFS: how to configure/indicate separate resource set(s) for Set A   Proposal 4-4: For UE-sided model inference, support to configure one resource set with resources in more than one measurement time instance for configuration of set B in BM Case 2.  Proposal 4-5: For UE-sided model inference, support to configure the periodicity and offset for set A, or configure the time gap between last measurement time instance of set B and predicted time instance of set A to indicate the time domain location of predicted time instance in BM Case 2. |
| Kyocera [35] | **Proposal 1** For a NW-sided AI/ML model for BM-Case1 and BM-Case2, for the configuration of inference results, RAN1 should consider the following options:   * Set A: No configuration for Set A is required for the UE, as the inference process is conducted transparently from the UE. * The gNB should configure reporting-related parameters, including the report size and the number of beams that the UE should report. * Explicit configuration of Set B is necessary, with resources for Set B explicitly defined for the UE.   **Proposal 2:** For the UE-side AI/ML model, particularly for BM Case-1, CSI-ResourceConfig could be used for configuring the resources for Sets A and B. The following scenarios should be further studied:   * + Alt1&4 can be deprioritized in cases where Set A does not need to be configured for either inference or performance monitoring.   + Sets A and B are configured using the same CSI-ResourceConfigId but different CSI-ResourceSetId.   + Sets A and B are configured using different CSI-ResourceConfigId.   **Proposal 3:** For UE-side AI/ML models, the association provides the UE with information about the beams that belong to Sets A and B. To facilitate this association, a IE is required. This IE could be defined within the CSI-ResourceConfigId or within a new AI/ML specific configuration. |
| LGE[36] | Proposal #6: Support Alt 1 or Alt 4 for Set A and Set B configuration for UE-sided AI/ML model.  Proposal #7: Regarding Alt 4 for Set A and Set B configuration,   * One or more separate resource set(s) for Set A can be configured outside of *CSI-ResourceConfig* * The set ID for Set A can be configured to each *CSI-ReportConfig*   Proposal #8: In order to support Alt 1 for Set A and Set B configuration, assistance information on relation/association between Set A beams and Set B beams should be provided to UE for the UE-side AI/ML model training and inference. 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 linear combining coefficients of Set B beams * Tx beam directions are represented as ordered numbers on a 2D or 3D coordinate |
| NEC[38] | Proposal 7: For triggering/initiating data collection at UE side for UE-side AI/ML model, support both  − Option 1: data collection initiated/triggered by configuration from NW, and  − Option 2: request from UE for data collection can serve as a supplementary. Moreover, Option 2 may need additional clarification on the steps that follow after request initiated by the UE.  Proposal 14: For configuration of Set A, separate CSI-ReportConfig for Set A and Set B are needed and a linkage is configured in CSI-ReportConfig for Set B to provide the Set A information.  Proposal 15: The configured resources for Set A should be available for other channel/signal (e.g., PDSCH) at least during model inference. |
| MTK [39] | Proposal 11: For the discussion of Set A/B configuration in CSI-ReportConfig, RAN1 makes a working assumption that Set A is assumed to be always configured (i.e., have corresponding resource ID/resource set ID, ResourceConfig ID), for the following reasons:   1. Set A is required for monitoring/data collection, and can be configured with longer periodicity 2. Set A is required for legacy UE   Proposal 12: For the design of csi-ReportConfig for UE-sided model, at least for BM Case-1, support Alt 2 and Alt 3 for Set A/B configuration  Proposal 13: For Alt2, Set A can be configured by the existing IE csi-RS-ResourceSetList. Set B can be configured by a new IE, in the form of either a bitmap, bitmap ID, a set of RS IDs, or ResourceSet ID  Proposal 14: For Alt3, Set A can be configured by the existing IE resourcesForChannelMeasurement in csi-ReportConfig. Set B can be configured by a new IE which has csi-ResourceConfigId as its value.  Proposal 15: For UE-side model inference BM-Case2, for periodic Set B RS resources, study the following alternatives for configuring the observation window length to UE:   * Alt1: observation window length is configured within LCM framework   + the length of observation window should be one of the conditions to BM functionalities * Alt2: observation window length is configured within CSI-framework   + Alt A: explicitly configured in csi-ReportConfig   + Alt B: implicitly configured in csi-ReportConfig     - For P/SP report: indicated through ReportPeriodicityAndOffset of the report and ResourcePeriodicityAndOffset of the associated RS resources of Set B     - For AP report: indicated by the triggered slot and report slot   Proposal 16: For UE-side model inference BM-Case2, for periodic Set B RS resources, study the following alternatives for configuring the prediction window periodicity/length to UE:   * Alt1: prediction window periodicity/length is configured within LCM framework   + the periodicity/length of prediction window should be one of the conditions to BM functionalities * Alt2: prediction window periodicity/length is configured within CSI-framework   + Opt 1: explicitly configured in csi-ReportConfig     - Applicable only to P/SP/AP reports   + Opt 2: implicitly configured in csi-ReportConfig     - Only applicable to P/SP reports   Proposal 17: For UE-side model inference BM-Case2, for aperiodic Set B RS resources, study the following alternatives for configuring the prediction window periodicity/length to UE:   * Alt1: prediction window periodicity/length is configured within LCM framework   + the periodicity/length of prediction window should be one of the conditions to BM functionalities * Alt2: prediction window periodicity/length is explicitly configured within CSI-framework   Proposal 18: For UE-side model inference, same design of report and resource configurations can be used for both BM Case1 and BM Case2.  Proposal 26: For UE-sided model, to save RS overhead, Set B is determined by NW from RAN1 perspective.  Proposal 27: For UE side data collection, NW needs to configure both Set A and Set B information to UE.  Proposal 28: For UE side data collection, for report and resource configurations, consider Set A and Set B resources are configured in one report configuration and reuse the following design alternatives of UE side model inference:   * Alt2: one CSI-ResourceConfigId is configured for both Set A and Set B * Alt3: two CSI-ResourceConfigId s are configured for Set A and Set B separately.   Proposal 29: RAN1 consider a common SetA/B resource configuration design for both UE side data collection and UE side model inference purposes, that is, Set A is configured in csi-ReportConfig::resourcesForChannelMeasurement; Set B is configured within/outside csi-ReportConfig::resourcesForChannelMeasurement  Proposal 30: To indicate whether a report configuration is for UE side model inference or UE side data collection, NW configures different reportQuantity for each purpose:   * reportQuantity = “predicted-CRI” or “predicted-CRI-RSRP”, UE measures Set B, and reports the predicted beam ID/RSRP among csi-ReportConfig::resourcesForChannelMeasurement (i.e., Set A) * reportQuantiy = “cri-RSRP” or “none” (existing values) but Set B is configured in csi-ReportConfig. UE reports as a legacy beam report, and perform data collection (i.e., model input is Set B measurements, model output is measurement of Set A (i.e., csi-ReportConfig::resourcesForChannelMeasurement) |
| Apple [40] | Proposal 3-4: data collection can be initiated/triggered by configuration from NW; or is requested from UE and then may be configured by NW at NW’s discretion. |
| KT[41] | Proposal 4. Further discuss the details based on Alt 1 and Alt 2 for configuration of Set A and Set B:   * Alt 1: one CSI-ResourceConfigId is configured for Set B * Alt 2: one CSI-ResourceConfigId is configured for Set A and Set B |
| Meta[42] | 1. **For configuration of set A for UE sided models, support Alt-1 and Alt-4, i.e., set A is not explicitly configured using CSIReportConfig.** |

### Issue #4.1: Configuration for RS for measurement of multiple time instances for BM-Case 2

##### (FL0) Proposal 4.1:

1. For UE-side AI/ML model inference, for BM-Case2, further study how NW can configure UE to report inference results of N future time slots, including

* NW- selected time instances: NW configure UE with N time instances relative to the reference time
* UE-selected time instances: NW configure UE to select and report N time instances relative to the reference time, with corresponding predictions
* Reference time definition: For example, MAC-CE/DCI activation slot, or UE reporting slot
* Maximum future time instance that can be predicted (e.g. 80ms/160ms/320ms/640ms/800ms/others)

|  |  |
| --- | --- |
| Supported companies | FL |
| Companies | View |
| FL | Valid point to be studied.  But details need to be considered together |
| NTT DOCOMO | We suggest focusing on NW-selected time instances. Since reported inference results are used for scheduling and beam management at gNB side, time instances of these reported inference results should be determined by gNB. |
| OPPO | Okay to further study the reporting of N future time slots.  One minor comment on reference time definition, it seems like a new definition in current spec, particularly for BM-Case2. To be more inclusive at current stage, could we also consider RRC configured reference time? The following change would look good to us now.   * Reference time definition: For example, RRC configured slot, MAC-CE/DCI activation slot, or UE reporting slot |
| HW/HiSi | 1. 1) From main bullet, it says NW to configure N, while in the 2nd sub-bullet, it says UE to select N time instances, which is contradictory with main bullet. If the intention is to enable UE flexibility of reporting dynamic number of N to NW, then better to reword the main bullet as in below. 2. 2) For the reference time, why do we consider a MAC CE or activation DCI as the reference? Each CSI report should have their own reference time, rather having a common reference time of such semi-static signaling.  Updated: (FL0) Proposal 4.1:  1. For UE-side AI/ML model inference, for BM-Case2, further study how ~~NW can configure UE~~ to determine the N future time slots as reported inference results ~~of~~, including  * NW- selected time instances: NW configure UE with N time instances relative to the reference time * UE-selected time instances: ~~NW configure~~ UE to select and report N time instances relative to the reference time based on NW configuration, with corresponding predictions * Reference time definition: For example, ~~MAC-CE/DCI activation slot,~~ CSI reference resource slot of the report, or UE reporting slot * Maximum future time instance that can be predicted (e.g. 80ms/160ms/320ms/640ms/800ms/others) |
| Fujitsu | What’s the exact difference between the NW-selected time instances and UE-selected time instances?   1. We prefer to the NW-selected time instances since this is for the inference reporting. But the UE should be able to report the preferred number of instances for prediction operation. |
| TCL | Support UE-selected time instances. This will facilitate the UEs to decide the reporting according to their capabilities. |
| Xiaomi | We prefer NW-selected time instances. As for UE-selected time instances, how does UE select?  As for the reference time definition, what does ‘MAC-CE/DCI activation slot’ mean? Does it mean that the reference time is the slot UE received the CSI report activation/triggering by MAC CE/DCI? If my understanding is right, we suggest to consider the last slot of the last history time instance, or the starting slot of the periodicity of predicted time instances as the reference time.   * Reference time definition: For example, the slot UE received CSI report activation/triggering by MAC-CE/DCI ~~activation slot~~, or ~~UE~~ CSI reporting slot, or the last slot of the last history measurement time instance, or the first slot of the periodicity of predicted time instances. * Maximum length of future time ~~instance~~ that can be predicted (e.g. 80ms/160ms/320ms/640ms/800ms/others) |
| ZTE | Fine to study the beam report configuration for future time instances as well as the reference time definition. However, referring to the R18 Doppler CSI principles, it shall be NW to configure the predicted time instances according to aligned model description information. Not sure the intention of UE-selected time instances in the second sub-bullet. |
| Qualcomm | All of what is mentioned in this proposal is based on UE capability report for BM-Case2. UE would report its capability about how long into the future it can predict (potentially accompanied by granularity), and all the other procedures mentioned in the proposal may follow after that capability report. Without such capability report, NW cannot arbitrarily ask for a certain future time instances to be predicted by UE. Also it should be mentioned that consistency across training and inference is assumed here for prediction. If UE has been trained to predict for every 40x (x = 1,2, 3, .., N), then configuring UE to report the next 10 msec may not be possible. |
| InterDigital | Do not support the proposal. We failed to understand the reason for selecting time instances especially by UE. Existing periodic/semi-persistent/aperiodic mechanism should be enough for reporting. |
| Ericsson | Support. Ok with the text clarifications above. |
| LG | Fine to study. |
| CATT | We prefer to support NW- selected time instances. Because the network requires performing beam management and scheduling decision based on reporting of UE, it’s better to let NW to select time instances rather than UE. And the motivation to introduce UE- selected time instances method is not clear.  Also, for reference time definition: For example, MAC-CE/DCI activation slot, CSI reference resource slot, measurement resource(s) related slot,or UE reporting slot |
| Nokia | We are not fully clear the need of NW selection or UE selection mentioning. Reference slot can not be dynamic by default. Such considerations are secondary and let’s first finalize the simple form of enabling this use-case.  Updated proposal   1. For UE-side AI/ML model inference, for BM-Case2, ~~further study how~~ NW can configure UE to report inference results of N future time slots, where future times slots refer to the slots after the slot that carrying the CSI reporting (i.e., reference time slot), including  * NW- selected time instances: NW configure UE with N time instances relative to the reference time * FFS: whether the UE can select N future slots (e.g., UE to select a reference time slot, and report N time instances relative to the reference time, with corresponding predictions) * Reference time definition: For example, MAC-CE/DCI activation slot, or UE reporting slot * Maximum future time instance that can be predicted (e.g. 80ms/160ms/320ms/640ms/800ms/others) |
| Panasonic | Since main bullet describes “how NW can configure UE to report inference results of N future time slots”, hence sub-bullets points should focus on a case of NW-selected time instance, i.e., NW configures UE with N time instances relative to the reference time.  For other case, it could be considered as FFS points. |
| MediaTek | We prefer NW-selected time instances. In terms of reference time definition, we suggest keeping the MAC-CE/DCI activation slot. Because the AI BM purpose is to substitute legacy measurement and reporting for a set of RS resources, we think the reference time for prediction should be related to the scheduled RS resources. For Aperiodic resources, the timing of the scheduled RS resources are defined by an offset to the DCI triggering slot. |
| Spreadtrum | Okay to further study the reporting of N future time slots. It should be mentioned that NW should first get the UE capability for how many future instances can be predicted. |
| CMCC | 1. We prefer to prioritize NW-selected time instances, as DOCOMO’s comment, result of beam prediction is used for scheduling, so gNB decides which time instance needs prediction result. 2. Regarding reference time definition, we support UE reporting slot. Referring to the configuration method of prediction instances for CSI prediction in R18 MIMO, UE reporting slot is clear and intuitive to determine the predicted instances combining time offset, N and time interval P of future time instances. For BM-case2 with large measurement window, compared with RRC slot or MAC-CE/DCI activation slot, time offset for UE reporting slot to determine the first predicted instance is smaller and has less overhead. |
| Kyocera | 1. We are supportive for the FL proposal |
| Sony | 1. For the configuration of N future time instances, it may depend on UE capability, channel status, and other factors. Therefore, we believe that further study is needed for both NW-selected and UE-selected time instances. |
| Sharp | 1. Support. It is important to define the specific time instance included in one beam report for BM Case 2; otherwise, it is not clear which results on a specific time instance should be included. Agree with other companies to propose removing MAC-CE/DCI activation slot. The term “MAC-CE/DCI activation slot” is not clear. Also, either configuring a time configuration or using a reference point with a specific time duration (window) to indicate the time instances works well, so we support both methods. |
| Futurewei | 1. Fine to study. We prefer NW-selected time instances. |
| NEC | We have a clarification question regarding UE-selected time instances: NW configure UE to select and report N time instances relative to the reference time, with corresponding predictions.   1. It looks like NW- selected time instances too, what exactly that UE can select for this option? |
| Lenovo | We prefer NW-select/configure N time instance. However, as HW pointed the sub-bullet and the main bullet are not aligned. We are fine with further study based on HW’s updated version. |
| Apple | HW’s update looks good. |

### Issue #4.2(on hold): Configuration of Set A and Set B

For UE-sided model at least for BM Case-1, *CSI-ReportConfig* is used for the configuration of inference results reporting

* FFS on the details in the *CSI-ReportConfig*, at least considering:
  + Alt 1: one *CSI-ResourceConfigId* is configured for Set B
    - FFS: how UE can determine the information about set A
    - Supported by: ZTE, Nokia, Fujitsu?,Samsung,xiaomi, LGE,KT,Meta
  + Alt 2: one *CSI-ResourceConfigId* is configured for both Set A and Set B
    - FFS: How to configure resource set(s) for Set A and Set B in *CSI-ResourceConfig*
    - Supported by: Futurewei, Spreadtrum, CMCC, Nokia(basic),Transsion, Sharp, Lenovo?, Fujitsu?, Interdigital, Qualcomm, Kyocera,MTK,KT
  + Alt 3: two *CSI-ResourceConfigId* s are configured for Set A and Set B separately
    - Supported by: Futurewei, CMCC, ZTE,OPPO, Huawei/HiSi, ITL, Transsion, Sharp, Panasonic, Lenovo, Fujitsu, DoCoMo, Samsung, Kyocera,MTK
  + Alt 4: one *CSI-ResourceConfigId* is configured for Set B, Set A is configured using separate resource set(s) other than that represented by *CSI-ResourceConfigId*
    - FFS: how to configure/indicate separate resource set(s) for Set A
    - Supported by: Intel, Nokia (open for optional), Sharp, DoCoMo,xiaomi, LGE,Meta
    - Postpone: vivo
  + Note: separate *CSI-ReportConfig* for Set A and Set B are not precluded.
  + Note: Not perform measurement for Set A and only perform measurement for Set B subject to the *CSI-ReportConfig*
  + FFS on the association between Set A and Set B with or without additional IE

Other necessary configuration are not precluded.

# Measurement report for NW-sided model

|  |
| --- |
| Agreement  For NW-sided model, for inference, in a beam report initiated by network, based on one measurement resource set, support the report of more than 4 beam related information in L1 signaling   * Note: Purpose, such as above “For NW-sided model, for inference”, will not be specified in RAN 1 specifications * FFS on the report content for beam related information * FFS on max number of reported beam related information in one report   Agreement  For NW-sided model, for inference report, at least for BM-Case 1, the content in a beam report in L1 signaling, support   * L1-RSRPs and corresponding beam information of Top M beam(s) with largest M measured value(s) of L1-RSRP(s) of a measurement resource set, where M is configured by gNB * If M = the size of the measurement resource set, the content is all L1-RSRPs and one beam index (i.e., CRI/SSBRI) for the largest measured value of L1-RSRP of a measurement resource set * FFS: L1-RSRPs and corresponding beam information of up to M beams within X dB gap to the largest measured value of L1-RSRP, X and M are configured by gNB, and whether/how to report number of reported beams * FFS on the maximum value of M (where M can be larger than 4) based on UE capability (M may or may not be different for different reporting contents) * FFS on beam information * Note: Purpose, such as above “For NW-sided model, for inference report, at least for BM-Case 1”, will not be specified in RAN 1 specifications   Agreement  At least for NW sided model, for the quantization of a reported L1-RSRP value at least for the report in L1 signaling, support   * Support differential L1-RSRP reporting with legacy quantization step and range   + FFS: larger quantization step(s) than the already supported legacy quantization step for differential L1-RSRP and/or for absolute L1-RSRP   + FFS: Smaller range(s) for differential L1-RSRP than the already supported legacy range |

### Summary from the contributions

|  |  |
| --- | --- |
| Companies | Proposals |
| Futurewei [1] | Proposal 2: For Rel-19 AI/ML-based BM, at least for inference for network-sided AI/ML model of BM-Case1, for the content in a beam report in L1 signaling,   * Support L1-RSRPs and corresponding beam information of up to M beams within X dB gap to the largest measured value of L1-RSRP, X and M are configured by gNB, and the number of reported beams is indicated in the beam report. * The beam information is CRI/SSBRI. * Increase the maximum value M from the existing “4” to “8” as a starting point. |
| Spreadtrum [2] | Proposal 3：For data collection for training at NW-side, the content in a beam report in L1 signaling, support   * L1-RSRPs and corresponding beam information of M beam(s), where M is configured by gNB   + If M = the size of the measurement resource set, the content is all L1-RSRPs and one beam index (i.e., CRI/SSBRI) for the largest measured value of L1-RSRP of a measurement resource set.   Proposal 4：For NW-sided model, both L1 and high-layer signaling can be used for data collection for training.  Proposal 5: For BM-Case1 and BM-Case2 with a network-side AI/ML model, larger quantization step(s) should not be considered at least for model inference.  Proposal 6: At least for NW-side model, for reported beam information, at least support Opt 0.   * Opt 0: legacy CRI/SSBRI, (i.e., index of resource in a resource set)   Proposal 9: Reporting multiple past time instances in one reporting instance for BM-Case2 is not needed.  Proposal 10: For BM-Case2, implicit report of time information should be supported. |
| GOOGLE[3] | Proposal 2: Support configurable quantization step and range for different L1-RSRP and absolute L1-RSRP report.  Proposal 14: For beam report for NW model training for temporal beam prediction, UE reports the following information:   * SSBRI/CRI with the best L1-RSRP measured for each of the measured slots configured by the NW * L1-RSRP for a subset of SSBs/CSI-RSs configured by the NW for each of the measured slots configured by the NW * Hypothetical measurement error for the subset of SSBs/CSI-RSs with L1-RSRP reported   Proposal 15: For beam report for NW model inference for temporal beam prediction, UE reports the following information:   * L1-RSRP for a set of SSBs/CSI-RSs configured by the NW for each of the measured slots configured by the NW * Hypothetical measurement error for the subset of SSBs/CSI-RSs with L1-RSRP reported   Proposal 16: For beam report for NW model monitoring for temporal beam prediction, UE reports the following information:   * SSBRI/CRI and L1-RSRP for the best SSB/CSI-RS from the SSBs/CSI-RSs configured as CMR for one or multiple slots configured by the NW * L1-RSRP for a subset of SSBs/CSI-RSs configured by the NW |
| Tejas[4] | Proposal 1: For inference report at NW sided model at least for BM-case 1, consider L1-RSRPs and corresponding beam information of up to M beams within X dB gap to the largest measured value of L1-RSRP.  Proposal 5: At least for NW-sided model, for differential L1-RSRP reporting, support for introducing a larger quantization step size for differential L1-RSRP reporting. |
| CMCC[5] | Proposal 1: L1 signaling is supported for NW-sided training data collection.  Proposal 2: For NW-sided model, for data collection for training, at least for BM-Case1, option 1 and 2 are supported for the report contents:   * Opt 1: L1-RSRPs from RS resources from one set of beams configured to UE   + for higher layer report, beam information is CRI of RS with largest L1-RSRP   + for L1 signaling, beam information can be bitmap * Opt 2: L1-RSRPs from RS resources from one set of beams configured to UE, and beam information of Top K from RS resources from another set of beams   + K is configured by RRC, value of K is 1~10   + for higher layer report, beam information is CRI of RS with largest L1-RSRP in measurement set and CRI of Top K RS in prediction set   + for L1 signaling, beam information is bitmap and CRI of Top K RS in prediction set   Proposal 3: If L1 signaling is applied for NW-sided training data collection, further discuss reporting all or a subset of L1-RSRPs from the resource set, at least including data omission, e.g., L1-RSRP(s) higher than a threshold.  Proposal 17: For NW-sided model, for inference report, at least for BM-Case 1, if the content in a beam report in L1 signaling is L1-RSRPs and corresponding beam information of Top M beam(s) with largest M measured value(s) of L1-RSRP(s) and M < the size of the measurement resource set, beam information is bitmap.  Proposal 18: For NW-sided model, for inference report, at least for BM-Case 1, the content in a beam report in L1 signaling, support:   * L1-RSRPs and corresponding beam information of beams within X dB gap to the largest measured value of L1-RSRP, X is configured by gNB.   + beam information is bitmap   + UL resource allocation for variable number of reported beams needs further discussion   Proposal 19: For NW-sided model, for inference report for BM-Case 2, temporal compression for overhead reduction is supported at least for aperiodic report. Beam information can be bitmap. |
| Intel [6] | It is not necessary to configure multiple resource sets associated to one L1 beam report for a network-sided model.  For a network-side AI/ML model, for BM-Case 1/2, for inference, the network may configure the size of the L1 report based on the measurements from set B. The configuration can be explicit or implicit based on measurement resources configured to the UE.  The number of beams to be reported in a single reporting instance may depend on the type of data being collected i.e., it may be configured to be different for training, inference and model monitoring.  The container for UE beam reporting should be dependent on the type of collected data i.e., inference/model monitoring data collection should use L1 based reporting, while data collection for training should use RRC or MAC-CE based reporting.  For content for data collection for training for NW-sided model, for BM-Case 1 per instance, support Opt. 1: L1-RSRPs from RS resources from one ~~or two~~ set~~s~~ of beams configured to UE.   * The beam information for Top K beams (corresponding to set A of beams, but transparent to the UE) can be determined implicitly from the reported L1-RSRP values for set of beams measured. * The set of beams configured to the UE can be the union of sets A and B of beams (transparent to the UE).   For content for data collection for training for NW-sided model, for BM-Case 2, support Opt. 1: One or multiple sets of L1-RSRPs for one or multiple time instances, respectively, from RS resources from one ~~or two~~ set~~s~~ of beams configured to UE.   * The beam information for Top K beams (corresponding to set A of beams, but transparent to the UE) can be determined implicitly from the reported L1-RSRP values for set of beams measured. * The set of beams configured to the UE can be the union of sets A and B of beams (transparent to the UE). * Regarding information on the time instances, study the following approaches including combinations thereof:   + implicit determination of time instances for multiple measurements based on configuration of two time windows that are defined in terms of the respective window length and time gap between them.   + explicit indication of time information for the time instances via reporting of timestamps associated with each measurement.   For a network-side AI/ML model, for BM-Case 2, for measurement reporting for inference (purpose to be transparent to a UE), the network may configure the UE with an observation window and the number of measurements on reference resources related to set B to be reported.   * The UE reports one or multiple L1-RSRP measurement results for the beams in the set B corresponding to one or multiple time instances, respectively, in a report with related timestamps of the measurements.   Proposal 17: For a network-side AI/ML model, for BM-Case 2, the UE may not need to be configured with a prediction window. |
| ZTE[7] | Proposal 1: For the contents of collected data from UE to NW, all three options can be supported to serve different LCM operations at the NW side.   * Opt.1: M1 L1-RSRPs (corresponding to M1 beams) with the indication of beams based on the measurement corresponding to a beam set, where M1 can be larger than 4, if applicable * 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 * Opt.3: M3 beam indices based on the measurement corresponding to a beam set, where M3 can be larger than 4, if applicable   Proposal 2: Regarding measurement results report, if measurement results of partial beams in a measured beam set are to be reported, support enhanced method (e.g., bitmap) for the indication of beam ID in UE reporting.  Proposal 7: For NW-sided model, for inference report, at least for BM-Case 1, the content in a beam report in L1 signaling, support L1-RSRPs and corresponding beam information of up to M beams within X dB gap to the largest measured value of L1-RSRP, X and M are configured by gNB.  Proposal 8: For the reporting of L1-RSRPs and corresponding beam information of up to M beams within X dB gap to the largest measured value of L1-RSRP, the legacy two parts encoding method for CSI reports should be reused as much as possible.  Proposal 9: For overhead reduction, support to specify threshold based beam reporting method with configurable minimum and maximum number of reported beam related information in a single report.  Proposal 10: For overhead reduction, support specification enhancements for data omission among samples (e.g., according to data quality).  Proposal 11: Support to specify the bitmap based method for the reporting of beam information given the fact that it can greatly reduce the reporting overhead in typical settings without compromising any beam prediction performance.  Proposal 12: At least for NW sided model, for the quantization of a reported L1-RSRP value, support differential L1-RSRP reporting with larger quantization step size applicable to the differential L1-RSRP.  Proposal 13: At least support L1 signaling for NW-side data collection irrespective of the purpose of data collection (e.g., model training, model inference, and performance monitoring)  Proposal 14: For NW-side model inference, the report content for beam related information comprises beam ID information and L1-RSRP, where the beam ID information can be reported by new beam indexing formats (e.g., bitmap) to reduce the reporting overhead.  Proposal 15: For NW-side model inference, the maximum number of reported beam related information in one report can be configured by the NW based on UE capability indication.  Proposal 16: Support enhancements to report information about measurements of multiple past time instances in one reporting instance for BM-Case2, at least from the following aspects.   * Indication of the timestamp information * Indication of the reference beam * Indication of the common beam information, e.g., a common/super set of beam IDs to be reported   Proposal 17: Support a common framework design for the measurement reporting of multiple past time instances for NW-side model and prediction reporting of multiple future time instances for UE-side model. |
| Ericsson [8] | Proposal 18 For NW-sided model inference/training, regarding max number of reported beam related information in one report, use 256 beams as a starting point, RAN1 studies the following mechanisms to support such report,   * Opt1) Multiple-resource sets, e.g. 64 RS in each set and UE report measurements for all RSs * Opt2) Extended resource sets, e.g. resource set supports 256 RS, and UE report measurements for all RSs  1. Proposal 19: For NW-sided model, regarding the content for data collection for training, conclude that there is no fundamental specification impact except extending the max number of reported beam related information in UE one report.    * Note that possible signaling overhead reduction might require specification impact. 2. Proposal 20: For NW-sided model inference, support NW configuration for UEs to pre-process set B beams to reduce reporting overhead, via:    * Support configuring reporting of only beams within X dB of the strongest beam,    * Support configuring reporting of at most N strongest set B beams. 3. Proposal 21: For NW-sided model inference, support methods for UEs to compress the set B temporal domain measurement results to reduce the reporting overhead.   Proposal 22 For NW-sided data collection, RAN1 studies possible “omission/selection of collected data” by the following aspects as a starting point,   * Possibility for UE to avoid signalling “duplicated” samples, * Possibility for UE to avoid signalling data based on certain events, one event can comprise that the UE experienced large channel variation during set A measurements. * Note: RAN2 can use such study when designing data collection procedures   Proposal 23 For NW-sided data collection, conclude that it is up to RAN2 on whether RRC/MDT procedures should be supported |
| Vivo [9] | Proposal 33: For data collection, for NW-side model, support to report UE measurement results via L1-layer signaling and higher-layer signaling.  Proposal 34: For data collection, for NW-side model, confirming the agreement from SI phase that more than 4 beams can be reported in a beam report.   The maximum number of reported beam related information in one report is related to UE’s capability.  Proposal 35: For data collection, for NW-side model, it is crucial to investigate approaches to minimize the overhead of the report transmitted through L1-layer or higher-layer signaling.  Proposal 36: For data collection, for NW-side model, report content supported in current specification can be re-used for BM-Case 1 and BM-Case 2, i.e. L1-RSRPs and corresponding beam information.  Proposal 37: For data collection, for NW-side model, enhancement on overhead reduction should be supported for BM-Case 1.  Proposal 38: For data collection and inference, for NW-side model, enhancement on overhead reduction should be supported for BM-Case 2, e.g. consider to use time domain data compression to reduce overhead.  Proposal 39: For inference, for NW-side model, if M = the size of the measurement resource set, report the L1-RSRP in a predetermined order corresponding to the measurement resource set.  Proposal 40: For inference, for NW-side model, support to report L1-RSRPs and corresponding beam information of up to M beams within X dB gap to the largest measured value of L1-RSRP, as well as the number of reported beams  Proposal 41: For inference, for NW-side model, support to introduce a beam index type indicator to specify the meaning of resource indicator, which indicates the reported beams are the beams with highest quality or lowest quality.  Proposal 42: For inference, for NW-side model, support to report of measurement results of multiple past time instances in one report for BM-Case 2.  Proposal 43: For data collection and inference, for NW-side model, support enhancements on quantization range and quantization step to reduce overhead for measurement results report.  Proposal 44: For data collection and inference, for NW-side model, support to report TRI (time resource indicator) instead of direct predicted beam resource indication scheme with implicit time stamp. TRI indicates where each of the unique reported beams locates in the future time occasions. |
| OPPO [10] | Proposal 2: For BM-Case2 with NW-side model, UE reports multiple measurement instances in a single beam reporting instance.  Proposal 5: For NW-side model training, UE reports the following contents to NW  • L1-RSRPs measurements of fixed Set B as model inputs  • Top-K L1-RSRP(s) and Top-1 Tx beam index as labels  Proposal 6: For BM-Case2 with NW-side model, the temporal domain information of collected data for training could be reported in an implicit manner, i.e. no explicit time stamps needed.  Proposal 7: Continue to discuss the content for data collection for NW-side model training without mentioning for higher layer or physical layer report. |
| Fujitsu [11] | Proposal 1:   * For training data collection, at least the following information should be included:   + Reference signal ID   + Beam quality, e.g., L1-RSRP   Proposal 2:   * For training data collection, the model input data and the ground truth data should be included. Whether and how to map/associate the model input data with the ground truth data could be further discussed per sub-use case for beam management.   Proposal 3:   * Regarding training data collection for NW-side model, the UE could report all the L1-RSRP for all the beams in Set A if Set B is subset of Set A. And the UE could report all the L1-RSRP for all the beams in Set B and Set A if Set B is different from Set A.   Proposal 4:   * For training data collection, RAN1 to further discuss the quantization for the model input data and ground truth data. High-resolution quantization and non-differential RSRP could be considered.   Proposal 12:   * For BM Case-1 with UE side model, if Set B is subset of Set A, whether predicted RSRP or measured RSRP should be applied for Set B beams for reporting of the inference results could be subject to AI/ML model performance. If the performance is not good, the measured RSRP is used, otherwise the predicted RSRP is used.   Proposal 19:   * For BM Case-1 with NW-side model, regarding the reporting content for Set B measurement results for inference operation, the beam related information could include CRI/SSBRI.   Proposal 27:   * For BM Case-2 with NW-side model, regarding the reporting content for Set B measurement results, the beam related information could include CRI/SSBRI and corresponding L1-RSRP.   Proposal 28:   * For BM Case-2 with NW-side model, RAN1 to discuss that the UE can report the preferred pattern for measurement including the number of measurement instances and the measurement interval. |
| CATT [12] | Proposal 18: For NW-sided model, for inference report, at least for BM-Case 1, support L1-RSRPs and corresponding beam information of up to M beams within X dB gap to the largest measured value of L1-RSRP, X and M are configured by gNB.  Proposal 19: For NW-sided model, for inference report, the maximum value of M based on UE capability should be discussed in the UE feature.  Proposal 20: For NW-sided model, the following options can be considered for the reported beam information   * Opt 1: Legacy CRI/SSBRI of a resource set, and resource set id if multiple resource sets consists set B; * Opt 2：The indicator for largest measured value of L1-RSRP, and a bitmap indicating RS index within a resource set, and resource set id if multiple resource sets consists set B.   Proposal 21: For NW-sided model for BM case-2, for inference, support to report largest L1-RSRP from N time instances and other differential L1-RSRP of N time instances within the measurement window in a pre-defined order in a beam report.  Proposal 22: At least for NW-sided model, for inference, support to introduce a larger quantization step size for differential L1-RSRP reporting. |
| TCL [13] | Proposal 7: The following enhancement on the beam management report should be considered.   * Alt. 1 Indicating the model type and/or a bitmap to indicate the selected report quantities. * Alt. 2 Indicating the model type and/or the report type to indicate the selected report quantities.   Proposal 8: The report for AI/ML BP may include L1-RSRP and/or post processed RSRP.  Proposal 9: RAN1 should consider the following enhancement on the report of AI/ML beam management.   * For overhead reduction purpose, study the quantization of report quantities, starting from the enhancement on the RSRP quantization. * Study the two-stage report mechanism using both PUCCH and PUSCH.   Proposal 10: For BM-Case2, the following overhead reduction approach can be considered.   * The report may be split into multiple groups for latency and overhead reduction, FFS the splitting rule and collision control mechanism. * The selection of predicted beams in the report can be indicated by a reference beam plus a bitmap indicating the appearance of predicted beams within the neighbourhood. |
| Lenovo [14] | Proposal 14: 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.  Proposal 15: Study schemes on differential RSRP report for UCI overhead reduction for larger number of beam reports in a beam report. |
| InterDigital [16] | Proposal 8: Support an overhead reduction mechanism for reporting data collection for inference (e.g., based on channel conditions).  Proposal 9: Support reporting of L1-RSRPs and corresponding beam information of up to M beams within X dB gap to the largest measured value of L1-RSRP.  Proposal 10: Support a pattern-based reporting mechanism for data collection for inference wherein the UE reports a pattern ID(s) and corresponding beam measurements of a subset of beams in Set B.  Proposal 11: Information about the time stamp for measurement instances should be supported.  Proposal 12: Reporting prediction results of multiple future time instances in one report should be supported. Beam inference reporting periodicity should be aligned with CSI-Reporting periodicity as a baseline.  Proposal 13: Support flexible configuration of multiple future time instance reporting (e.g., based on channel conditions, gNB configuration).  Proposal 14: Support both  ‘Option A: Predicted RSRP’  ‘Option B: Predicted RSRP, if the beam is not configured for corresponding measurement, and measured L1-RSRP if the beam is configured for corresponding measurement.’  Additionally, support a UE based selection procedure to report according to Option A or Option B.  Proposal 15: For network sided model, support enhanced UE reporting to report up to 64 RSRP values for whole Set A over multiple time instances.   * No CRIs/SSBRIs are reported and implicit beam indexes (e.g., by association with RSs and reported RSRPs) are used. * Information on measured past instances (e.g., time stamp) is supported. |
| Panasonic [17] | Proposal 1: NW-sided model inference, support to that a measurement window can be configured with the measurement resource set.  Proposal 11: Group-based beam reporting can be enhanced to support performance monitoring for NW-sided model. |
| Hyundai [18] | Proposal #1   * Discuss whether and how to arrange the order of Top-M beam reports considering their priority (e.g., based on measured RSRP) for inference for NW-sided model. |
| Nokia [19] | Proposal 18: Consider the following for a CSI report that enables beam prediction at the NW,   * For BM-Case1 and BM-Case2, consider enhancements for L1-RSRP quantization, increasing the differential L1-RSRPs in the report to X dB quantization step.   + FFS: To reduce UCI reporting overhead, discuss value(s) of X (e.g. X=3 and X=4 larger than legacy X>2dB) configurable to the UE.   Proposal 20: For BM-Case2, consider the following for a CSI report that enables beam prediction at the NW,   * Support common framework design when reporting beam related information for both BM-Case1 and BM-Case2. * Consider enhancements to report multiple past time instances in one reporting instance   + FFS: Number of consecutive measurements of beams/RSs to be made between reporting instances. * FFS: whether time stamp information can be derived from report based on report configuration.   Proposal 23: For BM-Case1 and BM-Case2, support NW-sided data collection (for training) for beam prediction related CSI reporting, and further discuss the following options,   * Option 1: Use CSI report to report L1-RSRPs for all resources in one or two measurement resource sets.   + NW assumptions on Set B and Set A combinations may decide whether it is one or two sets. * Option 2: Use CSI report to report L1-RSRPs for all resources in a first measurement resource set, and Top-1/K beam information in a second measurement set. * FFS: how to indicate assumption on Rx beams in the CSI report.   + “best” or “Quasi-optimal” Rx beam should be selected by the UE and reflected in the measurement reports. |
| Ruijie [20] | Proposal 5: Support larger quantization step(s) than the already supported legacy quantization step for differential L1-RSRP and/or for absolute L1-RSRP.  Proposal 6: Do not support smaller range(s) for differential L1-RSRP than the already supported legacy range.  Proposal 7: For NW-sided model, for inference report, at least for BM-Case 1, the content in a beam report in L1 signaling, for beam information it is proposed to include at least legacy beam index (i.e., CRI/SSBRI), or a bitmap which indicating RS index of a measurement resource set for overhead reduction purpose. |
| Samsung [21] | Proposal 1. For NW-side AI/ML model data collection for training, support at least the following as data collection content:   * L1-RSRP(s) for all beam(s) of Set B * L1-RSRP(s) for all beam(s) of Set A * Top-K Beam ID(s) for Set A * Related timestamp * The information to facilitate model training (FFS details)   Proposal 2. For NW-side AI/ML model data collection for training, at least support the enhancement to use high layer signaling to convey data collection content.  Proposal 3. For NW-side AI/ML model inference, for CSI-ReportConfig with the measurements for more than 4 beams in one reporting instance   * Support differential L1-RSRP reporting with larger quantization step(s) and/or smaller range(s) * Consider to reduce the overhead of beam report using two-part CSI   Proposal 4. For NW-side AI/ML model inference, for CSI-ReportConfig with the measurements for more than 4 beams in one reporting instance, consider the applicability of the following report quantity:   * 'cri-RSRP', 'ssb-Index-RSRP', 'cri-RSRP-Index', 'ssb-Index-RSRP-Index'.   Proposal 5. For NW-side AI/ML model inference, support CSI-ReportConfig with measurements for each of multiple past time instances in one reporting instance.   * FFS: How to identify the multiple past time instances * FFS: The support of P/SP/AP reporting |
| Transsion [22] | Proposal 1: Regarding data collection for NW-side AI/ML model, support Opt.1 and Opt.2,   * Opt 1(w omission): L1-RSRPs and corresponding beam information of Top M beam(s); * Opt 2 (w/o omission): All L1-RSRPs.   Proposal 2: Regarding the signaling for training data collection, support higher-layer signalling to report the contents of training data.  Proposal 4: Regarding the quantization of a reported L1-RSRP value, support larger quantization step(s) than the legacy quantization step for differential L1-RSRP and/or for absolute L1-RSRP. |
| ETRI [23] | Proposal 10: For the NW-sided model, reducing measurement overhead is necessary for model inference, especially in the case of temporal domain prediction.  Proposal 11: For the NW-side model, support the method of omitting RSRP values based on differences in RSRP values during model inference.  Proposal 12: Support methods for reducing UE reporting overhead during data collection for training when the AI/ML model is located on the NW-side.   * The NW limits the maximum number of L1-RSRP values that the UE can transmit through CSI reporting. * The NW specifies an L1-RSRP threshold, requiring the UE to report the L1-RSRP values of Tx beams that exceed this threshold. |
| CAICT [24] | Proposal 5: for NW-sided model report, the introducing of X dB gap could be deprioritized.  Proposal 6: For the content for data collection for NW-sided model training, L1-RSRPs from RS resources from one or two sets of beams configured to UE should be supported for high layer signaling report.  Proposal 7: Finer step size should be considered for differential L1-RSRP feedback for NW-sided model report. |
| DOCOMO [25] | Proposal 18: Consider overhead reduction for more than 4 beam related information in L1 signaling.   * Large quantization step size for Set B measurement reporting * Reporting of measurements from multiple time instances in one reporting instance. |
| Sharp [26] | Proposal 6： For NW-side model, for inference report, at least for BM-Case 1, support configuring more than 4 measurement results (e.g., L1-RSRP) in one CSI report as a UE capability.  Proposal 7： For NW-side model, for inference report, at least for BM-Case 1, when the size of the measurement resource set is configured for reporting, except for the largest measured value of L1-RSRP, the order of differential L1-RSRP values in the report can be based on the ascending order of beam index.  Proposal 8： For NW-side model, for inference report, at least for BM-Case 1, beam information of Top M beam(s) in a beam report support:   Option 1: legacy M CRI/SSBRI fields where the M CRI/SSBRI fields indicate Top M beams.   Option 2: a legacy CRI/SSBRI field and a bitmap where one CRI/SSBRI filed indicates a beam index with largest measured L1-RSRP value and the bitmap indicates remaining M-1 beams.   The choice between option 1 and option 2 can be configured by gNB.  Proposal 9 ：For NW-side model, for BM-Case 2, at least for aperiodic reporting:   The set B can be consistuted by one or more aperiodic CSI-RS resource sets, triggered by a single triggering state.   Support of reporting measurement results from multiple past time instances for a single report configuration.   Support of enabling the triggering of an aperiodic CSI-RS resource set for channel measurement on multiple past time instances for a single report configuration. |
| Fraunhofer HHI, Fraunhofer IIS [29] | Proposal 11: Prior to the selection of a container for data collection, study the AI/ML purposes separately in terms of the amount of data to be reported, acceptable latency, security and reliability.  Proposal 12: Support L3 measurements as a container for L1-RSRPs reporting for training/re-training purposes given its increased payload size, relaxed latency requirement and higher reliability based on the usage of error-correction mechanisms.  Proposal 16: The use of a predicted beam that is not measured/received by the UE for beam indication is supported. |
| ITL [31] | Proposal 1: Regarding data collection for NW-side AI/ML model, support the report of more than 4 beam related information similar with that of inference.  Proposal 2: It is proposed that both Option 1 and Option 2 can be considered for the contents on data collection for NW side model training.  Proposal 3: For NW-side model data collection for training, it is proposed to at least support the higher layer signaling to convey data collection contents.  Proposal 4: It is proposed to determine the number of possible beam-related information by referencing the size of that content and the capacity of the possible container after concluding the discussion on which report content is supported.  Proposal 5: It is beneficial to use larger quantization step(s) than the already supported legacy quantization step(s) for L1-RSRP.  Proposal 8: For NW-sided AI/ML model inference, it is proposed to use CRI and SSBRI for the beam information of Top M beams with largest M measure values(s) of L1-RSRP(s) in a beam report in L1 signaling. The ordering CRI and SSBRI can be based on the values of L1-RSRPs of Top M beam(s).  Proposal 9: For NW-sided AI/ML model inference, it can be additionally supported to configure reporting of only beams within X dB gap of the strongest beam.  Proposal 10: For NW-sided AI/ML model inference for BM-case-2, support to report largest L1-RSRP from N time instances and other differential L1-RSRP of N time instances in a predefined order or time stamp-based method in a beam report.  Proposal 11: For NW-sided model inference, the max number of M (e.g. up to 256) reported beam related information in one report can be configured by the NW based on UE capability signaling. It does not need to be varied based on other reporting content.  Proposal 14: For BM-Case2 with the NW-sided model, it should be discussed on the details with how measurements from multiple time instances can be reported with less overhead. |
| KDDI [32] | Proposal 1: For the L1-RSRP reporting of the NW-sided model in BM-Case 2, consider supporting the reporting of the difference over time.  Proposal 2: An option with a quantization step size larger than 2 dB and with a wider range should be supported.  Proposal 3: Support reporting of L1-RSRPs and corresponding beam information of up to M beams within X dB gap to the largest measured value of L1-RSRP, where X and M are configured by gNB.  Proposal 4: The UE should be able to report the number of beams it intends to report, either periodically or based on network-specified events. |
| Huawei/Hisi[33] | Proposal 1: For training/monitoring data collection for both NW-side model and UE-side model, RAN1 to discuss the potential mechanism to enable the UE to perform CSI measurements on larger sizes of beam set(s), including:   * Alt 1: The beam set(s) for measurement consist of multiple resource sets each with legacy size (up to 64) of resources. * Alt 2: The beam set(s) for measurement consist of one resource set with increased size of resources, e.g., up to 256.   Proposal 2: For the maximum number of reported beam related information in one report,   * The max number of reported beams can be the same as the number of beams in the measured beam set(s), e.g., up to 256. * Consider BM-Case 1 as a starting point. For BM-Case 2, RAN1 needs to decide on whether or not to include multiple past time instances in one report.   Proposal 3: From RAN1 perspective, for NW-side model, for monitoring and/or training, conclude that L1 signaling can also be used.   * Note: The conclusion can be interpreted that the agreement from RAN1#116 for the report of more than 4 beam related information in L1 signaling does not need to be limited for inference. * Note: Purpose, such as above “For NW-side model, for monitoring and/or training”, will not be specified in RAN1.   Proposal 4: For report of data collection in higher layer for NW-side model for BM-Case 1 and BM-Case 2, each data sample includes L1-RSRPs from RS resources from one set of beams configured to UE for per time instance.   * FFS on whether/how any corresponding beam information needs to be reported explicitly or not. * Note 1: For BM-Case 2, the beam information and/or L1-RSRP information over multiple time instances can be derived from multiple data samples. * Note 2: Purpose, such as above “For NW-side model for BM-Case 1, BM-Case 2”, will not be specified from RAN1 perspective.   Proposal 5: For measurement report for NW-side AI/ML model, support an adaptive number of beams based on threshold, i.e., L1-RSRPs and corresponding beam information of up to M beams within X dB gap to the largest measured L1-RSRP value, X and M are configured by gNB, with M>4.  Proposal 6: For the report of measured L1-RSRPs for the NW-side model, reuse the legacy quantization step and range for L1-RSRP, i.e., no need to introduce larger quantization steps or smaller ranges for L1-RSRP.  Proposal 7: For the report of beam information for the NW-side model, for Opt 1(w omission): L1-RSRPs and corresponding beam information of Top M beam(s) of a resource set, support at least Opt A: CRI/SSBRI,   * For Opt B: bitmap and beam index of strongest RSRP, it can be considered if the needed payload size is smaller than Opt A with respect to the resource set size and number of reported beams.   Proposal 18: For reporting overhead reduction of NW-side AI/ML model, regarding data compression under BM-Case 2, the necessity of reporting measurements across multiple past time instances in one report is not clear, considering the larger latency and impact on gNB scheduling it would inflict. |
| Xiaomi [34] | Proposal 3-2: For data collection of NW-side AI/ML model training, regarding to reported beam information, the beam index with largest measured value of L1-RSRP should be reported for differential L1-RSRP reporting.  Proposal 3-3: For data collection of NW-side AI/ML model training, regarding to reported beam information, support UE to indicate the RS IDs whose L1-RSRP are not reported because of lower than threshold to reduce overhead.  Proposal 3-6: If one CSI-ReportConfig is used for set B and set A configuration for data collection for NW-side AI/ML model training, consider to support more than one reportquantity in one CSI-ReportConfig. |
| Kyocera [35] | Proposal 5: For training data collection of a NW side AI/ML model, in BM Cases 1 and 2, RAN1 should further study the following:  • Support using the CSI framework for inputs and labels collection of data samples for the purpose of training the AI/ML model. New IEs may be introduced, e.g., “CSI-AIML-ResourceSet” and “AIML-info-RSRP” to provide the gNB the flexibility to define any arbitrary Set B.  • Support reporting all Set B measurements during the data collection and training phase.  Proposal 6: For NW-sided model, for inference report, at least for BM-Case 1, the content in a beam report in L1 signaling, support:   * L1-RSRPs and corresponding beam information of Top M beam(s) with largest M measured value(s) of L1-RSRP(s) of a measurement resource set, where M is configured by gNB * If M = the size of the measurement resource set, the content is all L1-RSRPs and one beam index (i.e., CRI/SSBRI) for the largest measured value of L1-RSRP of a measurement resource set * FFS: L1-RSRPs and corresponding beam information of up to M beams within X dB gap to the largest measured value of L1-RSRP, X and M are configured by gNB, and whether/how to report number of reported beams * For the maximum value of M, and in order to support the cases when Set B is equal to Set A, support reporting the size of Set A (256 beams) as a starting point. * For the beam information, support reporting CRI/SSBRI as a starting point.   Note: Purpose, such as above “For NW-sided model, for inference report, at least for BM-Case 1”, will not be specified in RAN 1 specification |
| LGE [36] | Proposal #1: For NW-sided AI/ML in temporal DL Tx beam prediction, support the following UE reporting enhancements for data collection:   * Past/present best N beam(s) per time stamp * Tendency/variance of best N beam(s)   Proposal #4. Regarding the report of more than 4 beam related information in L1 signaling for NW-sided model, support CRI/SSBRI for beam information as legacy, e.g., CRI/SSBRI(s) of Top M beam(s) and the corresponding M L1-RSRP(s) are reported. |
| NVIDIA [37] | Proposal 4: For BM-Case 2, at least introduce specification support for using historical optimal beam index and/or L1-RSPR measurement based on Set B of beams as AI/ML model input. |
| NEC[38] | 1. Observation 1: The max number of reported beam in one report should be not less than the number of resources in the Set B, and it is related to the number of historical time instances allowed in one report.   Proposal 11: For overhead reduction for BM-Case2, support a larger RSRP quantization step size for the historical results obtained earlier, e.g., with a longer time from measurement to model inference.  Proposal 12: For a variable Set B selected from a set of pre-configured Set B patterns, indication related to the selected Set B (e.g., index of a group of beams) needs to be reported. For a variable Set B that is a subset of measured beams Set C, criterions/thresholds for determining the Set B need to be defined.  Proposal 13: For BM-Case2, at least measurement window should be configured for periodic beam report. |
| MTK [39] | Proposal 23: For NW side data collection, support higher layer signaling as baseline.  Proposal 24: For content for data collection for training, for both UE-sided model and NW-sided model, at least for BM-Case 1 per instance, further study the following options:  Note: the intention is for higher layer report   * Opt 1: L1-RSRPs from RS resources from one or two sets of beams configured to UE * Opt 2: L1-RSRPs from RS resources from one set of beams configured to UE, and beam information of Top K from RS resources from another set of beams |
| APPLE [40] | Proposal 3-3: L1 beam reporting for performance monitoring for NW-side model is supported.  Proposal 4-1: to control feedback overhead, beam reporting for BM Case-1 consists of   * Indication of the strongest beam index * The strongest beam’s RSRP * Bitmap to indicate un-omitted beams * Differential RSRPs for uno-omitted beams except the strongest beam * Indication of the number of un-omitted beams   Proposal 4-2: to control feedback overhead, beam reporting for BM Case-2 consists of   * Indication of the strongest beam index among all occasions * Bitmap to indicate un-omitted/omitted beams   + Alt. 1: bitmap size equals to the number of set B beams across occasions   + Alt. 2: bitmap size equals to the number of set B beams at a single occasion * The strongest beam’s RSRP * Differential RSRPs for un-omitted beams except the strongest beam * Indication of the number of un-omitted beams   Proposal 4-3: to control feedback overhead for the UE-side model, beam reporting for BM Case-2 consists of   * Indicating a subset containing top beams across time instances * Indicating a bitmap of selected top beams at time instances, the bitmap is over the cardinality of the subset by the number of future time instances.   Proposal 5-1: for AI/ML beam management, the effective time for beam reporting has the CSI reference source as the reference instead of beam report time. |
| Meta [42] | Proposal 3: For BM Case 2 measurement report, multiple time instances with a configured number of beams per time instance should be supported and the absolute value of time instances can be derived by NW based on periodicity of measurement RS transmissions.  Proposal 4: For training data collection for NW sided model, support L1-RSRP reporting for a configured set of beams using higher layers. |

### Issue #5.1: Measurement report for BM-Case2

##### (FL0) Proposal 5.1:

For NW-sided model, for BM-Case 2, support to report of measurement results of multiple past time instances in one report, [at least for inference].

* Differential L1-RSRP reporting over multiple time instances is used
* FFS on how to configure the time instances to the UE
* FFS whether include the time stamp information explicitly or implicitly
* Note: Purpose, such as above “For NW-sided model, for BM-Case 2, [at least for inference]”, will not be specified in RAN 1 specifications

|  |  |
| --- | --- |
| Supported companies | FL |
| Companies | View |
| FL | Majority companies support the above proposals. There is clear benefit. Hope we can agree on this direction and move on to details |
| OPPO | Support the FL proposal. |
| HW/HiSi | Do not support.  In our view the legacy mechanism is not only sufficient, but also better suited than the feature suggested in this proposals. As soon as the gNB receives a CSI report for a time instance, it has the flexibility to either perform spatial/temporal domain BM immediately or to perform temporal domain BM when the measurements for subsequent time instances of an observation window are all received. This has ower latency compared to if the gNB would have to wait until all the measurements of multiple past time instances of the observation window have been received, it will cause larger latency.  Another concern that we have are potential gNB scheduling restrictions, at least if the report is triggered by DCI. Since out-of-order scheduling is an error case, no other PUSCHs could be scheduled before the CSI report, and also the time delay between the DCI and the CSI report could become longer than what is supported by existing K2 values. |
| Fujitsu | Generally fine with the proposal. |
| TCL | Agree |
| Xiaomi | We would like to clarify the scenario to report of measurements results of multiple past time instances in one report.  For example, history measurement time instance is T1, 2T1, 3T1 and 4T1, and the predicted time instance is 5T1. If UE report the measurement results of T1, 2T1, 3T1 and 4T1 after 4T1, it means the measurement results of T1, 2T1, 3T1 is useless for gNB to determine the best beam of T1, 2T1, 3T1. Is it the case that the best beam of T1, 2T1, 3T1 is predicted by gNB before? in this case, if set B= set A, there is no RS overhead reduction. So can we clarify that this proposal is only used for set B is a subset of set A and set B is different from Set A? |
| ZTE | General fine. Since only the reporting of RSRP is mentioned in the proposal, we suggest to add another sub-bullet to clarify the reporting of beam information which is deemed necessary. For example, based on temporal correlations, the beam information of a common beam subset consisting of Top-K beams of all time instances is reported, and then the beam information specific to each time instance is selected only from the common beam subset.   * FFS on how to report the beam information |
| Qualcomm | For the baseline assumption (assuming we do not have the above proposed enhancement) UE has already reported the Top-K beams from Set B in previous reporting instances (as agreed before) and the main question here is what the delta is with that baseline in mind, which would justify this proposal. Maybe we can have further discussion on this prior to agreeing. |
| InterDigital | Fine |
| LG | Support. |
| CATT | Support. |
| vivo | Another issue is the report of beam information. At least beam information shall be supported as a content, which is similar as BM Case 1. Then whether and how to compress the overhead of beam information over multiple time instances.  Hence we suggest to add one more sub-bullet.   * The report of measurement results includes beam information of the multiple time instances. * FFS whether and how to compress the overhead of beam information over multiple time instances. |
| Fraunhofer | Agree. |
| Nokia | OK |
| KDDI | Support |
| Panasonic | Support |
| MediaTek | We can analysis this proposal under the following two scenarios:   * + - 1. Set B = Set A: for NW-side model, all the measured Set A L1-RSRP at each time instance needs to be reported. The overhead is huge regardless of whether UE reports separately for each time instance or in one report for all the time instances. The feasibility of this use case for NW-side model needs to be discussed before we discuss the spec change. **We need to justify first if the benefit of NW-side BM Case2** can mitigate the overhead of reporting these huge number of measurements.       2. Set B is a subset/different from of Set A: for this use case, at each observation time instances, we assume legacy BM is used, i.e., there is another report configured for UE to perform legacy beam measurement on Set A (UE reports only the best 4 beam + L1-RSRP). We don’t see the benefit of reporting Set B at each observation time instance with the legacy report. Therefore, **we support to report multiple instances in one report for this case**   Also, we would like to note that it is more reasonable and beneficial to use higher layer signaling for NW-side BM-Case2 inference reporting, especially if multiple time instances of measurements are sent in one report. Note that during SI, the minimum periodicity of measurement used by all the companies that we’ve seen is 40ms. This periodicity is even longer than the report periodicity of immediate MDT in RAN2. Therefore, we think higher layer signaling is more suitable. We suggest the following changes of wording:  For NW-sided model, for BM-Case 2, when Set B is a subset/different from of Set A, support to report of measurement results of multiple past time instances in one report, [at least for inference].   * Differential L1-RSRP reporting over multiple time instances is used * Note: the feasibility of NW-sided model, BM-Case 2 when Set A =Set B, needs to be justified * FFS on L1 or higher layer signalling for this report * FFS on how to configure the time instances to the UE * FFS whether include the time stamp information explicitly or implicitly * Note: Purpose, such as above “For NW-sided model, for BM-Case 2, [at least for inference]”, will not be specified in RAN 1 specifications |
| Spreadtrum | We don’t see strong motivation to support include multiple past instances measurement results in one report. Including multiple past measurements in a single report can result in significant overhead but no improvement in AI performance. If we report each time instance measurement, gNB can perform some analysis (e.g. model selection) based on the measurement results received first. Thus, reporting multiple past time instances in one reporting instance for BM-Case2 is not needed.. |
| CMCC | Fine with the proposal. |
| Kyocera | Support the FL proposal |
| Sony | We support the proposal. |
| Sharp | Support. |
| Futurewei | Not support. As mentioned by Qualcomm, the baseline already works and it is unclear about the benefit of the proposal, let alone the potentially added latency and restriction on NW scheduling. |
| NEC | We would like to understand if this proposal ‘Differential L1-RSRP reporting over multiple time instances is used’ means that the reference RSRP is the largest RSRP over multiple time instances. |
| Lenovo | Support |
| Apple | Support the FL’s proposal. |

### Issue #5.2: Data collection for training for NW sided model

##### (FL0) Proposal 5.2:

Conclusion

For data collection for training for NW-sided model, from RAN 1 perspective,

* L1 signaling can be used
* It is up to RAN2 on whether to use RRC/MDT procedures

RAN 1 will further study on whether/how to further reduce the reporting overhead for data collection for training other than for other purposes(inference/monitoring) assuming L1 signaling and potentially higher layer signaling.

|  |  |
| --- | --- |
| Supported companies | FL |
| Companies | View |
| FL | At least the first sentence as conclusion reflects the fact and should be able to be agreed.  The second sentence also reflects the fact with many proposals from companies.  In addition, I think it may be good if we can send RAN 2 LS on there decision on whether they will use RRC/MDT. |
| NTT DOCOMO | Support. |
| OPPO | For NW-side training data collection, we think higher-layer signaling should be prioritized over L1 signaling. It is not time critical which can be fully facilitated by MDT-based data collection.  In our view, what RAN1 should do is to determine the content for data collection as what Ran1 did in last RAN1#117 meeting. |
| HW/HiSi | Support the principle, but similar to other agreements for NW-side model, the purpose should not be specified and a corresponding note should be added. Updated (FL0) Proposal 5.2: Conclusion  For data collection for training for NW-sided model, from RAN 1 perspective,   * L1 signaling can be used * It is up to RAN2 on whether to use RRC/MDT procedures   RAN 1 will further study on whether/how to further reduce the reporting overhead for data collection for training other than for other purposes(inference/monitoring) assuming L1 signaling and potentially higher layer signaling.  Note: Purpose, such as above “For data collection for training for NW-sided model”, will not be specified in RAN 1 specifications |
| Fujitsu | Don’t support the proposal. RAN2 has already agreed to use higher layer signaling for training data collection for NW-side model. |
| TCL | Support both L1 and higher layer signaling |
| Xiaomi | Support |
| Qualcomm | The main question with regards to supporting L1 signaling for NW-side training is not feasibility, rather its necessity. Whether it is necessary to report L1 signaling for the purpose of NW-side training or not, which in our view the necessity is not justified given the needed latency aspects. Additionally, RAN2 is separately working on designing a framework for data collection for NW-side training regardless of what happens in RAN1, so the reason for duplicate signaling efforts should be elaborated first. |
| InterDigital | Support both L1 and higher layer signaling |
| Ericsson | Support |
| LG | Not support. If RAN2 specifies L2 based solution for data collection for training (e.g., RRC/MDT procedures), it does not make sense to specify L1 based solution “additionally” which would result in two different features for a same functionality. |
| CATT | Support. |
| Fraunhofer | Regardless of higher layer signaling, L1 signaling can be used for any purpose, hence, training is also not precluded. |
| Nokia | OK. It is not important to list second bullet. |
| MediaTek | Agree with Fujitsu and OPPO. We suggest giving Potential proposal 5.3 another try in this meeting. |
| Spreadtrum | Support. |
| CMCC | Support. |
| Kyocera | Support the FL proposal |
| Sharp | Support. And also OK with the Note proposed by Huawei |
| Futurewei | Support FL’s proposal. |
| NEC | Support |
| Lenovo | Support |
| Apple | Some UEs capable of helping NW-side model inference and performing performance monitoring may not support data collection for training. So this should be made as a separate UE capability, and consequently the L1 signaling if supported can be different from the L1 signaling for support inference/performance monitoring. |

##### (FL1) Proposal 5.2a+Proposal 5.5a+Proposal 5.6a:

**Conclusion:**

For data collection for training for NW-sided model, from RAN 1 perspective,

* L1 signaling can be used
* It is up to RAN2 on whether to use RRC/MDT procedures

~~RAN 1 will further study on whether/how to further reduce the reporting overhead for data collection for training other than for other purposes(inference/monitoring) assuming L1 signaling and potentially higher layer signaling, including:~~

**Proposal:**

For NW-sided model, at least for training data collection, at least for BM-Case 1,

* for the content in a beam report in L1 signaling,
  + support L1-RSRPs and corresponding beam information of up to M beams within X dB gap to the largest measured value of L1-RSRP, where X and M are configured by gNB
    - FFS: whether/how to report number of reported beams
* for the quantization of a reported L1-RSRP value, support
  + Y dB (FFS: Y=3 and/or 4) quantization step(s) than the already supported legacy quantization step for differential L1-RSRP
  + Note: smaller range(s) and larger quantization step(s) for absolute L1-RSRP in differential L1-RSRP reporting are not precluded and can be studied separately.

Note: Purpose, such as above “For NW-sided model, at least for training data collection, at least for BM-Case 1”, will not be specified in RAN 1 specifications

|  |  |
| --- | --- |
| Companies | View |
| FL | L1 signaling can be used. But for training data collection purpose, optimization should be supported to reduce the overhead.  For proposal 5.5 and proposal 5.6, some company questions about the motivation of this. If this package of the proposal cannot fly.  The next step will be: to discuss the content of data collection via RRC. Please don’t comment on it can be done via L1!!!! |
| CMCC | Fine with L1 signaling and overhead reduction with RSRP threshold, the determination of new quantization step value can be postponed. |

### Issue #5.3(on hold): Data collection for training for NW sided model for BM-Case1

Potential proposal 5.3:

For content for data collection for training for NW-sided model, at least for BM-Case 1 per instance, further study the following options:

Note: the intention is for higher layer report

* Opt 1: L1-RSRPs from RS resources from one set [FFS: two sets] of beams configured to UE
  + FFS on whether/how the corresponding beam information needs to be reported explicitly or implicitly.
  + Note: overhead reduction is not precluded.
* Opt 2: L1-RSRPs from RS resources from one set of beams configured to UE, and beam information of Top K from RS resources from another set of beams
  + K is configurable. FFS: K values
  + FFS on whether/how the corresponding beam information needs to be reported explicitly or implicitly.
  + Note: overhead reduction is not precluded.
* FFS: how to define UE capability, [e.g., the max number of RS resources, the max number of L1-RSRPs, the max beam information of Top K]

### Issue #5.4(on hold): Data collection for training for NW sided model for BM-Case2

Potential proposal 5.3:

For content for data collection for training for NW-sided model, for BM-Case 2, further study the following options:

Note: the intention is for higher layer report

* Opt 1: one or multiple sets of L1-RSRPs corresponding to one or multiple time instances, where L1-RSRPs in each set are from one [or multiple] resource set(s) for each time instance
  + The resource set(s) for each time instance can be same or different
  + All and a subset of L1-RSRPs from each corresponding resource set is supported.
  + FFS on whether/how the corresponding beam information needs to be reported explicitly or not.
  + FFS on whether/how the corresponding time instance information needs to be reported explicitly or not.
* Opt 2: one or multiple sets of L1-RSRPs corresponding to one or multiple time instances, and one or multiple set of beam information of Top K corresponding to other one or multiple time instances
  + where L1-RSRPs in each set are from one [or multiple] resource set(s) for each time instance
  + where beam information of Top K in each set are from one [or multiple] resource set(s) for each time instance
  + The resource set(s) for each time instance can be same or different
  + All and a subset of L1-RSRPs from each corresponding resource set is supported.
  + FFS on whether/how the corresponding beam information to L1-RSRPs need to be reported explicitly or not.
  + FFS on whether/how the corresponding time instance information needs to be reported explicitly or not.
* FFS on how to determinate a subset of L1-RSRPs, including
  + Alt 1: Top M beam(s) is the beams with largest M measured values of L1-RSRPs, where M is configured by gNB
  + Alt 2: All beams within X dB gap to the largest measured value of L1-RSRP
* Differential L1-RSRP reporting is supported,
  + FFS on whether to support differential L1-RSRP per time instance or across multiple time instances

### Issue #5.5: Quantization and L1-RSRP omission

##### (FL0) Proposal 5.5:

At least for NW sided model, for the quantization of a reported L1-RSRP value at least for the report in L1 signaling, support

* Y dB (FFS: Y=3 and/or 4) quantization step(s) than the already supported legacy quantization step for differential L1-RSRP
* Note: smaller range(s) and larger quantization step(s) for absolute L1-RSRP in differential L1-RSRP reporting are not precluded and can be studied separately.

|  |  |
| --- | --- |
| Supported companies | FL, NTT DOCOMO,New H3C |
| Companies | View |
| FL |  |
| OPPO | It seems pre-mature to determine the quantization steps, e.g. 3 or 4 dB.  Note that the L1-RSRP difference can be used as performance metric for monitoring. If RAN1 decide the L1-RSRP difference with value 3 or 4 dB can trigger an event, then the quantization error with 3 or 4 dB can bring similar impact over performance monitoring, then it would confuse either NW or UE on LCM decision. |
| HW/HiSi | Not support.  We have concern on the applicability of this proposal.   * For inference, the L1-RSRP report can be used for multiple purposes in a UE transparent manner, even including legacy operations. Changing the quantization level to a coarser value may impact the performance of these operations (such as scheduling, or pairing) in a negative manner. Also, when Set B is different from Set A, it may be difficult to identify the proper narrow beam direction, since the same quantization level may correspond to a multitude of beams. * For training, the overhead is not a large issue anyway since the training data collection is non-real time. As the sole purpose of changing the quantization granularity and range would be an optimization for overhead reduction, this enhancement may not be essential for training data collection either.   If studied further, this proposal should be discussed together with other overhead reduction techniques, e.g. reporting an adaptive number of beams. |
| Fujitsu | Similar view as OPPO. |
| TCL | Agree |
| Xiaomi | Is it for inference or other purpose? If it is for performance monitoring or training with L1-RSRP of set A, we share same concern as OPPO. |
| ZTE | For the main bullet, it should be the measured RSRP value (instead of the reported RSRP value) that needs to be quantized and then reported to the NW.  For the first sub-bullet, we prefer to only keep 4dB for Y. 4dB quantization resolution for the reporting of differential RSRP is more generally be used by companies during R18 evaluations. Besides, with Y=4dB, the differential RSRP can be quantized to a 3-bit value and one bit can be saved for the reporting of each differential RSRP. While for 3dB quantization resolution, there's no such benefit i.e., still 4 bits are needed.  For the last note, the benefit of enhancements for the reporting of absolute RSRP is marginal and evaluation verification is deficient.  Therefore, we suggest the following revisions.  At least for NW sided model, for the quantization of a ~~reported~~ measured L1-RSRP value at least for the report in L1 signaling, support   * Y dB (FFS: Y=~~3 and/or~~ 4) quantization step(s) than the already supported legacy quantization step for differential L1-RSRP   Note: smaller range(s) and larger quantization step(s) for absolute L1-RSRP in differential L1-RSRP reporting are not precluded and can be studied separately. |
| Qualcomm | Lower priority issue at this stage, given that the legacy quantization step has been supported. Whether to support this or not depends heavily on the number of beams from Set B that are reported and exactly how much overhead saving this proposal would amount to, which would guide us into whether we should support such enhancement or not. |
| InterDigital | Do not support |
| Ericsson | Not support. The current quantization is sufficient, there are other overhead reduction techniques that are more important. |
| LG | Do not support. We don’t think there is clear benefit of large quantization step. |
| CATT | We are generally ok with this proposal. And share same feeling with OPPO, it’s premature to decide the value of the quantization step. |
| vivo | We are open to discuss this issue, which is an issue of trading off performance and overhead. Given for NW side model, a lot of reporting data is required, RSRP quantization overhead can be reduced for the relatively weaker RSRPs. |
| Fraunhofer | Agree. |
| Nokia | Not fully sure what exactly the new content on this compared to last meeting agreement as Y values listed as FFS. |
| KDDI | Support. |
| Spreadtrum | Same view as QC. Since the legacy quantization step has been supported, how much overhead can be saved should be further evaluated. |
| CMCC | Since the purpose is unknown to UE, how can UE distinguish reporting for training data collection and reporting for legacy BM? Prefer to discuss the necessity of larger quantization step before we discuss the detail. |
| Sharp | Support. And suggest removing the Y=3. Noted that the current L1 differential RSRP ranges from 0 to -30dB. If we follow the current value range and use the Y=3, we still require 4 bits, as same as the legacy 2dB step, to represent the 10 combination. |
| Futurewei | Not support. Legacy quantization step works well and we don’t see the need of adding new ones. |
| Lenovo | Do not support. We share similar view with OPPO. |
| Apple | Low priority |
| Kyocera | Similar view as OPPO |

(FL1) Proposal 5.5a:

At least for NW sided model, for the quantization of a reported L1-RSRP value at least for the report in L1 signaling for data collection, support

* Y dB (FFS: Y=3 and/or 4) quantization step(s) than the already supported legacy quantization step for differential L1-RSRP
* Note: smaller range(s) and larger quantization step(s) for absolute L1-RSRP in differential L1-RSRP reporting are not precluded and can be studied separately.

|  |  |
| --- | --- |
| Companies | View |
| FL | Most of companies support data collection via L1.  There is clear motivation for quantization step relaxation base on SI output.  Why not? |

### Issue #5.6: L1-RSRP omission

##### (FL0) Proposal 5.6:

For NW-sided model, for inference report, at least for BM-Case 1, the content in a beam report in L1 signaling, support

* Support L1-RSRPs and corresponding beam information of up to M beams within X dB gap to the largest measured value of L1-RSRP, where X and M are configured by gNB
  + FFS: whether/how to report number of reported beams

|  |  |
| --- | --- |
| Supported companies | FL |
| Companies | View |
| FL |  |
| NTT DOCOMO | Not support the proposal. It is unclear if the adaptive number of reported beams can reduce the overhead. If the adaptive number is adopted, the field indicating the number of beams needs to be reported additionally. Also, it requires CSI part 1 and CSI part 2 partition which has not been supported in beam measurement reporting. Considering the large specification impacts and unclear gain, we think the following already agreed beam reporting is sufficient for inference results reporting in NW side model.   * L1-RSRPs and corresponding beam information of Top M beam(s) with largest M measured value(s) of L1-RSRP(s) of a measurement resource set, where M is configured by gNB |
| HW/HiSi | Support |
| Fujitsu | Don’t support. Similar view as NTT DoCoMo.  The existing agreement is sufficient for NW-side model inference. We don’t see strong need for further optimization. |
| New H3C | OK in general |
| TCL | Agree |
| Xiaomi | Support |
| ZTE | Support. For threshold based measurement reporting, the reporting of the number of reported beams is necessary to determine the UCI payload size. Therefore, we suggest the following minor revision.   * + FFS: ~~whether/~~how to report number of reported beams |
| Qualcomm | Share similar view as Docomo. |
| InterDigital | Support. We believe that overhead reduction gain is clear and specification impact is not significant as part 1 and part 2 design was used for CSI from Rel-15. |
| Ericsson | Support |
| LG | Do not support. We have similar view with DOCOMO. |
| CATT | We support FL’s proposal. It is beneficial for overhead reduction. For example, the UE can adapt dynamically the number of beams in one report based on the measurement results. Compared with the reporting with fixed number of beams, it can avoid reporting redundant beams when channel state is poor and reporting insufficient beams when channel state is good. |
| vivo | Fine with the proposal |
| Nokia | Not sure about this proposal. Needs some discussion to understand the need. |
| KDDI | Support. We have similar view with InterDigital. |
| Panasonic | We share same view as DOCOMO. We cannot see motivation for this proposal. |
| MediaTek | Support |
| CMCC | Support. Overhead reduction with RSRP gap is also applicable to other purposes such as L1 signaling based training data collection or monitoring for NW side model. |
| NEC | Support. Since the number of reported beams may be insufficient to execute model inference, NW needs to configure a minimum number of M. |
| Lenovo | Support |
| Apple | Support |

(FL1) Proposal 5.6:

For NW-sided model, ~~for inference report,~~ at least for training data collection, at least for BM-Case 1, the content in a beam report in L1 signaling, support

* Support L1-RSRPs and corresponding beam information of up to M beams within X dB gap to the largest measured value of L1-RSRP, where X and M are configured by gNB
  + FFS: whether/how to report number of reported beams

|  |  |
| --- | --- |
| Companies | View |
| FL | Most of companies support data collection via L1.  There is clear motivation for quantization step relaxation base on SI output.  Why not? |

### Issue #5.7 Beam information

##### (FL0) Proposal 5.7:

For NW-sided model, for inference report, at least for BM-Case 1, the content in a beam report in L1 signaling, further study the following options for beam information when M<the size of measurement resource set (for both with or without XdB gap to the largest measured value of L1-RSRP):

* Option 1: CRI/SSBRI of a measurement resource set
* Option 2: a bit map to indicate the reported beams of a measurement resource set and one beam index (i.e., CRI/SSBRI) for the largest measured value of L1-RSRP of a measurement resource set

Note: Purpose, such as above “For NW-sided model, for inference report, at least for BM-Case 1”, will not be specified in RAN 1 specifications

|  |  |
| --- | --- |
| Supported companies | FL |
| Companies | View |
| FL | M=the size of the set has been agreed.  Let’s narrow down to limited options first. |
| NTT DOCOMO | Since at least for M≦4 the overhead issues do not exist and Option 1 is already supported, new enhancements, such as Option2, should be considered only for M>4. For this reason, we suggest the proposal focusing only when 4<M<the size of measurement resource set. |
| HW/HiSi | We support the direction of this proposal, but have some further comments.   * Option 1 is the legacy solution and already supported, it does not need to be studied anymore in our opinion. * Option 2 could be studied further. Depending on the configuration (size of the measured set and number of reported beams) either Option 1 or Option 2 are more overhead efficient. Depending on the configuration, Option 1 or Option 2 could be autonomously selected.  Updated (FL0) Proposal 5.7: For NW-sided model, for inference report, at least for BM-Case 1, the content in a beam report in L1 signaling, ~~further study~~ the following options for beam information when M<the size of measurement resource set (for both with or without XdB gap to the largest measured value of L1-RSRP) are considered:   * Option 1: CRI/SSBRI of a measurement resource set * FFS: Option 2: a bit map to indicate the reported beams of a measurement resource set and one beam index (i.e., CRI/SSBRI) for the largest measured value of L1-RSRP of a measurement resource set that is used if the needed payload size is smaller than Option 1 with respect to the resource set size and number of reported beams. . |
| Fujitsu | We think the legacy method can work, e.g., Option 1. And it is sufficient. |
| New H3C | OK in general |
| TCL | Agree |
| Xiaomi | Fine with Huawei’s update. |
| ZTE | Support |
| Qualcomm | OK in principle. However, we do not need “(for both with or without XdB gap to the largest measured value of L1-RSRP)”. The main purpose of XdB gap was to educe overhead. Now, if we have this enhancement for that scenario (which is still FFS, by the way) we will have a variable size bitmap, which has its own complications. Why do we need such dual optimization given the complications it entails? |
| InterDigital | Support Option 1 and do not support Option 2. |
| Ericsson | Not a critical proposal. We should first determine the max number of beams to be reported by the UE, then we could understand the need for Option 2. |
| LG | OK to study. Suggest to firstly discuss the maximum value of M to make the study clearer. The sentence within round brackets “(for both with or without XdB gap to the largest measured value of L1-RSRP)” is relevant to P5.6 thus needs to be removed. |
| CATT | OK |
| vivo | If the majority of beams are selected for reporting, the UE could report partial measured L1-RSRPs and the index of lowest-quality beams based on a beam set to the gNB, with the beam indices in the report used to indicate the lowest-quality beams.  Similarly, if the minority of beams are selected for reporting, the UE could report partial measured L1-RSRPs and the index of highest-quality beams based on a beam set to the gNB, with the beam indices in the report used to indicate the highest-quality beams as legacy behaviour  Hence from the motivation of reducing CRI/SSBRI overhead, we propose to add one sub-bullet under Option 1.   * Option 1: CRI/SSBRI of a measurement resource set   + FFS the definition of CRI/SSBRIs, e.g., whether they are the strongest or weakest beams in the resource set |
| Fraunhofer | Agree. |
| Nokia | We have discussed this issue for multiple meetings. Time to agree on a solution now. Both options have regions that have a lower overhead. Few updates to the proposal.  Updated proposal  For NW-sided model, for inference report, at least for BM-Case 1, the content in a beam report in L1 signaling, ~~further study~~ supports the following ~~options~~ for beam information when M<the size of the measurement resource set (N) ~~(for both with or without XdB gap to the largest measured value of L1-RSRP)~~:   * If M\*log2(N) <= N + log2(N), support Option 1: CRI/SSBRI of the measurement resource set * Otherwise, support Option 2: a bit map (dimension of N) to indicate the reported beams of the measurement resource set and one beam index (i.e., CRI/SSBRI) for the largest measured value of L1-RSRP of the measurement resource set   Note: Purpose, such as above “For NW-sided model, for inference report, at least for BM-Case 1”, will not be specified in RAN 1 specifications |
| Spreadtrum | Agree with Huawei’s update. |
| CMCC | There is a tradeoff between option1 and option2 when number of reported beams changes. Support UE to select between option1 and option2 depends on size of setA and number of reported beams. |
| Sharp | Support. Network can configure UE which option to be used, that is, which option would be an efficient solution in UCI reporting. |
| Futurewei | Option 1 works and we don’t see the need of adding Option 2. |
| NEC | We may need to wait whether multiple measurement resource sets can be configured, in that case, resource set ID may be also needed. |
| Lenovo | Option 1 has been specified for legacy beam report and we think it should be at least supported. We can further study whether option 2 can be further studied. |
| Apple | we support Option 2. Nokia’s version also looks fine. |
| Kyocera | Support |

##### (FL1) Proposal 5.7:

For NW-sided model, for inference report, at least for BM-Case 1, the content in a beam report in L1 signaling, ~~further study~~ the following options for beam information when M<the size of measurement resource set ~~(for both with or without XdB gap to the largest measured value of L1-RSRP)~~ are considered:

* Option 1: CRI/SSBRI of a measurement resource set
* FFS: Option 2: a bit map to indicate the reported beams of a measurement resource set and one beam index (i.e., CRI/SSBRI) for the largest measured value of L1-RSRP of a measurement resource set
  + Note: combination of option 1 and option 2 are not precluded and condition (if any) can be studied

Note: Purpose, such as above “For NW-sided model, for inference report, at least for BM-Case 1”, will not be specified in RAN 1 specifications

|  |  |
| --- | --- |
| Companies | View |
| FL | With or without change further study to considered. It is further study. |
| New H3C | OK in general |

# Inference result report for UE-sided model

|  |
| --- |
| Agreement  For UE-sided model, at least for BM-Case1, for content in the report of inference results, support   * Opt 1: Beam information on predicted Top K beam(s) among a set of beams * Opt 2: Beam information on predicted Top K beam(s) among a set of beams and RSRP of predicted Top K beam(s) among a set of beams * At least K=1 and more, FFS on max value * FFS on beam information * FFS on the definition of predicted Top K beam(s) * FFS on definition of reported RSRP when applicable * FFS on other information in the report with potential down selection among the following options * Opt 3: Beam information on predicted Top K beam(s) among a set of beams and probability information of predicted Top K beam(s) among a set of beams   + FFS on the quantization method of probability information   + Probability information is the probability of the beam to be the Top 1 or Top K beam * Opt 4: Beam information on predicted Top K beam(s) among a set of beams, RSRP of predicted Top K beam(s) among a set of beams, and confidence information of the RSRP   + FFS on definition of reported RSRP   + FFS on the definition and quantization method of confidence information * Other options are not precluded.   where the set of beams is Set A, i.e., the beams for UE prediction.  Agreement  For report content of inference results for UE-sided model for BM-Case 1, for the RSRP ofpredicted Top K beam(s) in the report of inference results, when applicable, further study the following options:   * Option A: Predicted RSRP * Option B: Predicted RSRP, if the beam is not configured for corresponding measurement, and measured L1-RSRP if the beam is configured for corresponding measurement * Where the predicted RSRP is based on AI/ML output * Note: Support both Option A and Option B is not precluded.   Agreement  For UE-side AI/ML model inference, for BM-Case2, support to report inference results of N(N>=1, FFS on N) future time instance(s) in one report   * wherein information of inference results of one time instance is as in one report for BM-Case 1   + Note: overhead reduction is not precluded * FFS on details   Agreement  For report content of inference results for UE-sided model for BM-Case 2, the RSRP ofpredicted beam(s) in the report of inference results, is the predicted RSRP, where the predicted RSRP is based on AI/ML output |

### Summary from the contributions

|  |  |
| --- | --- |
| Companies | Proposals |
| Futurewei [1] | Proposal 4: For Rel-19 AI/ML-based BM, for UE-sided model, at least for BM-Case1, for content in the report of inference results, do not support Opt 3 and Opt 4.   * Opt 3: Beam information on predicted Top K beam(s) among a set of beams and probability information of predicted Top K beam(s) among a set of beams * Opt 4: Beam information on predicted Top K beam(s) among a set of beams, RSRP of predicted Top K beam(s) among a set of beams, and confidence information of the RSRP   Proposal 11: For BM-Case2 with a UE-side AI/ML model, the number of reported beams can adaptively selected by UE for each N instance |
| GOOGLE[3] | Proposal 5: For beam report based on UE model inference for SD beam prediction, support the followings on the remaining open issues:   * Beam information is defined as a beam indicator (BI) indicating one of the beams from a configured codebook * The selection of the “top-K” beams are up to UE implementation   + Spec only defines the number of reported beams * The reported RSRP should be defined as the predicted RSRP based on a reference transmission power * Support both option 3 and option 4   Proposal 17: For beam report based on UE model inference for temporal beam prediction, UE reports the following information:   * N beam index(es) for one or multiple configured predicted slot(s) based on a configured beam codebook and measurement of a set of SSBs/CSI-RSs configured as CMR   Proposal 18: Support the UE to report prediction results for multiple time instances, i.e., N>1 is supported   * Support the UE to report the beam report for a subset of future time instances from the future time instances configured by the NW |
| CMCC[5] | Proposal 24: For UE-side AI/ML model inference, for BM-Case2, to report inference results of N future time instance(s) in one report   * The time interval of N future time instance(s) is P (P≥1) slots, P is configured by gNB * For the reference time to determine the earliest time instance from the N future time instance(s), option 1 is supported:   + Option 1: Based on the time domain resource for the report + offset     - the time instance(s) other than the earliest one is determined by slot of report + offset + N + P     - measured time instances and predicted time instances may have the same or different time interval   + Option 2: Based on the CSI reference resource corresponding to the report   + Option 3: Based on the transmission occasion of the CSI-RS/SSB resource in Set B for the report   Proposal 25: For report content of inference results for UE-sided model for BM-Case 1 whether the predicted L1-RSRP is reported can be configured by gNB, whether/how to differentiate measured L1-RSRP and predicted L1-RSRP needs further discussion. |
| Intel [6] | Proposal 16: For model inference of UE-side AI/ML model, for reporting predicted beams and related RSRP (Opt-2), differentiation between measured L1-RSRP and predicted RSRP from a model is needed. It can be based on an additional bit of information in the report per reported beam.  Proposal 17: Opt. 3 and Opt. 4 identified during RAN1 #116 towards defining contents of inference results reporting for UE-sided models are discussed in context of performance monitoring and discussions are not duplicated in context of contents of report with inference results.  Proposal 18: For a UE-side AI/ML model, for BM-Case 1/2, the number of best predicted beams may be configured to the UE by the network subject to UE capability. Accordingly, the L1 report from UE to gNB after inference may be defined based on the configured value. |
| ZTE[7] | Proposal 21: For UE-sided model, at least for BM-Case1, for content in the report of inference results, support the reporting of RSRP in Option 2 and probability information in Option 3 for beam selection at the NW side.   * The predicted Top-K beams are the K beams with the highest predicted RSRP (or probability information) * The beam information is CRI or SSBRI for BM-Case1, FFS for BM-Case2 * The reported RSRP is predicted RSRP, if the beam is not configured for corresponding measurement, and measured RSRP if the beam is configured for corresponding measurement, when applicable   Proposal 22: 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.  Proposal 23: For content in the report of inference results, at least for Opt 1: Beam information on predicted Top K beam(s) among a set of beams, the ranking information of Top K beams can be conveyed by reporting Top-K beam IDs in a descending order according to the predicted RSRP (or probability information) values. |
| Ericsson [8] | 1. Proposal 9 For UE-side AI/ML model inference, for BM-Case2, support that UE can update reported inference results of N future time instances after such report. 2. Proposal 10 For UE-sided model inference, regarding the FFS on beam information, conclude that such information at least comprises the CSI-RS resource indicator (CRI) and SSB resource indicator (SSBRI)   Proposal 11 For UE-sided model inference, regarding the FFS on potential down selection of option 3 and 4, support both option 3 and 4  Proposal 12 For UE-sided model inference, conclude that value of K could be adaptive and based on the UE-sided model output (e.g. support that UE report K beams so the probability of one of them being strongest is above a certain threshold) |
| Vivo [9] | 1. Proposal 25: For inference, for UE-side model, support to report predicted L1-RSRPs and corresponding beam information of up to M beams within X dB gap to the largest predicted value of L1-RSRP, as well as the number of reported beams. 2. Proposal 26: For inference, for UE-side model, when UE reports L1-RSRP of Top-k beams predicted from Set A, and for those that belongs to Set B, the UE reports the actually measured RSRP corresponding to the beam. 3. Proposal 27: For inference, for UE-side model, support time stamp information as beam content for BM-Case2. 4. Proposal 28: For inference, for UE-side model, support time domain compression of beam resource indication to further reduce report overhead with a report including results of multiple occasions. 5. Proposal 29: For inference, for UE-side model, support beam information on predicted Top K beam(s) among a set of beams and RSRP of predicted Top K beam(s) among a set of beams per predicted instance for content in the report for BM-Case 2. 6. Proposal 30: For inference, for UE-side model, support new report quantity to differentiate between AI-based predicted report with legacy measured report. 7. Proposal 31: For inference, for UE-side model, support to report predicted L1-RSRP of indicated beam(s), e.g. current beam indicated by TCI, with predicted L1-RSRP of top-k beams in one prediction report. 8. Proposal 32: For inference, for UE-side model, support to report TRI (time resource indicator) instead of direct predicted beam resource indication scheme with implicit time stamp. TRI indicates where each of the unique reported beams locates in the future time occasions. |
| OPPO [10] | Proposal 8: For UE-side model, support UE to report Alt.3 for inference, i.e. the beam information on predicted Top K beam(s) and probability information of predicted Top K beam(s).  Proposal 9: For UE-side model, clarify the beam information of predicted Top-K beam(s) as SSBRI/CRI associated with Set A.  Proposal 10: For BM-Case2 with UE-side model, support overhead reduction (e.g. enhanced differential L1-RSRP reporting) to facilitate multiple future time instances within one report.  Proposal 11: For BM-Case2 with UE-side model, the timestamp of future time instance(s) could be implicitly reported to NW. |
| Fujitsu [11] | Proposal 11:   * For BM Case-1 with UE side model, regarding the reporting contents of the inference results,   + Option 2 is preferred, i.e., Beam information on predicted Top K beam(s) among a set of beams and RSRP of predicted Top K beam(s) among a set of beams.   + The beam information could include CRI/SSBRI and CC ID. |
| CATT [12] | Proposal 10: For UE-sided model, at least for BM-Case1, for report content of inference results, the reported beam information can be the RS indicator(s) (e.g., legacy CRI/SSBRI) or the pre-defined beam index of the predicted Top K beam(s).  Proposal 11: For UE-sided model, at least for BM-Case1, for report content of inference results, the reported RSRP type of predicted Top K beam(s) can be configured by gNB with the following options:   * Option A: Predicted L1-RSRP; * Option B: Predicted RSRP, if the beam is not configured for corresponding measurement, and measured L1-RSRP if the beam is configured for corresponding measurement. |
| Lenovo [14] | Proposal 8: Considering the prediction window for BM-Case2 with the following modes:   * Mode 1: * Mode 2:   is the CSI reference resource for the beam report in slot n. and is a value configured by RRC. Which mode is used can be configured by RRC according to UE capability.  Proposal 9: For UCI overhead reduction on inference result reporting for BM-case2, to support that   * differential RSRP quantification with reference to the largest value among all future time instances * reporting unique beams of future time instance and corresponding time-stamp indicator. * an indication for informing whether to report all beams or unique beams.   Proposal 13: For a beam report associated with AI inference, the UE indicates that the reported beams are predicted beams or measured beams in the beam report. |
| Sony [15] | Proposal 1 : For BM-Case 2, the network can dynamically indicate the time window size for data collection and output to the UE separately, or the UE can dynamically report its preferred time window size to the network.  Proposal 2 : For BM-Case 2, the number of consecutive slots within each future time instance can be dynamically indicated by the NW, or the UE can dynamically report its preferred number of consecutive slots to the network.  Proposal 3 : For the UE-side model, considering the contents of the report of inference results, we support Options 1, 2, and 3. |
| Panasonic [17] | Proposal 6: To differentiate between prediction and measurement results, the following options can be considered:   * Option 1: To introduce prediction-related metrics as the reporting quantities. * Option 2: To introduce different resource sets in a report configuration for prediction and measurement. |
| Hyundai [18] | Proposal #2   * Discuss whether not only predicted RSRP but also measured L1-RSRP are considered to select Top K beams for report content of inference results for UE-sided model BM-Case1.   Observation #1   * The difference between measured L1-RSRP and predicted L1-RSRP can be beneficial for NW to monitor channel condition and UE’s inference.   Proposal #3   * Discuss whether/how UE reports difference between measured L1-RSRP and predicted RSRP or assistance information if the beam is configured for corresponding measurement instead of reporting measured L1. |
| Nokia [19] | Proposal 1: For BM-Case1, consider the following for a CSI report that enables beam prediction at the UE,   * For the supported Opt.1 and Opt.2,   + K = 1, 2, and 4. K is configurable to the UE.   + Beam information refer to CRIs corresponding to Set A   + RSRP of predicted Top K beam(s) reported similar to legacy L1-RSRP reporting.   + For the case of Set B is subset of Set A, RSRP of predicted Top K beam(s) can be corresponding to the measured or predicted L1-RSRP, and NW can determine it based on corresponding CRI. * Study further details on Opt. 3, where reporting content shall be “Top-K Predicted-CRIs, probability info” corresponding to a Set A.   + K = 1, 2, and 4. K is configurable to the UE.   + Beam information shall refer to CRIs corresponding to Set A   + Probability information shall be the probability of the beam to be the Top 1.   + FFS: how to report probability information (e.g., quantization method of probability information of predicted Top K beam(s)). * Do not support Opt.4. * FFS: whether measured beam related quantities (CRIs, L1-RSRP) of Set B can be configured to report within the same beam report.   Proposal 2: For BM-Case2, support reporting of “Top-K Predicted-CRIs for N time periods” or “Top-K Predicted-CRIs, predicted L1-RSRPs for N time periods” corresponding to a Set A, where K and N are configurable to UE.   * K = 1, 2, and 4 * N = 1, 2, 3, and 4 * FFS: Other overhead reduction options to apply when K and N values are large, including changes to the reporting format and details of periodic/aperiodic CSI reports. |
| Samsung [21] | Proposal 9. For UE-sided model, at least for BM-Case1, for the definition of the beam information on predicted Top K beam(s) among Set A, consider the following options:   * Option 1. The beam information is predicted SSBRI/CRI. * Option 2. The beam information is predicted beam indicator.   Proposal 10. For report content of inference results for UE-sided model for BM-Case 1, for the RSRP of predicted Top K beam(s) in the report of inference results, support both Option A and Option B.  Proposal 11. For UE-sided model, at least for BM-Case1, for content in the report of inference results, support:   * Opt 3: Beam information on predicted Top K beam(s) among a set of beams and probability information of predicted Top K beam(s) among a set of beams   + FFS on the quantization method of probability information   + Probability information is the probability of the beam to be the Top 1 or Top K beam   Proposal 12. For UE-sided model, at least for BM-Case1, for content in the report of inference results, further study the method to convey the ranking information of the predicted Top K beams in case of K > 1.  Proposal 13. For UE-sided model inference, for the reporting of predicted Top K beam(s) among Set A,   * K is configured by CSI-ReportConfig and the maximum configurable value of K is subject to UE capability.   Proposal 14. For UE-side AI/ML model inference, for BM-Case2, to report inference results of N future time instance(s) in one report   * Each of the N future time instance(s) consists of P (P≥1) consecutive slots   + FFS: How to determine P * For the reference time to determine the earliest time instance from the N future time instance(s), consider the following options:   + Option 1: Based on the time domain resource for the report   + Option 2: Based on the CSI reference resource corresponding to the report   + Option 3: Based on the transmission occasion of the CSI-RS/SSB resource in Set B for the report   + FFS: whether the above options are also applicable to the time instance(s) other than the earliest one   + FFS: If N>1, whether the time domain separation between two adjacent time instance(s) from the N future time instances are the same * FFS: How to define measurement window(s) for the inference results of the N future time instance(s)   Proposal 15. For UE-side AI/ML model inference, support differential L1-RSRP reporting for predicted beams. |
| ETRI [23] | Proposal 13: For the UE-side model, support the CSI report format for temporal domain beam prediction, including the optimal K beam information along with RSRP information from multiple time instances.  Proposal 14: For the UE-sided model, support the following differential L1-RSRP methods for temporal domain beam prediction.   * Relative to a single absolute RSRP value at each time point. * Relative to the absolute RSRP values at the first time point.   Proposal 15: Support the indication field for model inference when an AI/ML model is located on the UE-side.  Proposal 16: Support the CSI report format with only the predicted beam indices without L1-RSRP for model inference when an AI/ML model is located on the UE-side. |
| DOCOMO [25] | Proposal 7: Beam information on predicted top K beam(s) should be represented by CRI/SSBRI to follow the existing specification.  Proposal 8: Since the probability of predicted top K beam(s) and the confidence of predicted RSRP represent the performance metric of beam prediction, the necessity of these information should be discussed in performance monitoring not in inference result reporting.  Proposal 9: Reported time instance should be determined based on the parameters indicating the interval between two future time instances and the offset between the last referred transmission occasion and the first future time instance for AI/ML.  Proposal 11: For BM-case2, UE reports only if receiving multiple CSI-RS transmission occasions no later than CSI reference resource after CSI report (re)configuration  Proposal 12: Support the following payload overhead reduction on the reporting of predicted results at multiple time instances.  ・For the combination of CRI/SSBRI and time instance achieving the largest RSRP, absolute RSRP representation  ・For the remaining combinations of CRI/SSBRI and time instance, differential RSRP representation from the above absolute RSRP  Proposal 13: Enhancements of CSI processing units should be considered for beam prediction. |
| Qualcomm [27] | Proposal 8 For UE-side beam prediction, for content in the report of inference results, regarding FFS on potential down-selection among Option 3 and Option 4, support Option 4. |
| Fraunhofer HHI, Fraunhofer IIS [29] | Proposal 14: For UE-sided models, for inference, examine whether and how to report confidence of predictions.  Proposal 13: For BM-Case2, for inference, the reference time of the predicted time instances shall be determined based on the CSI reference resource or the transmission occasion of Set B resources. |
| ITL [31] | Proposal 18: For UE-sided model, beam information for predicted beam (e.g., model ID and CRI, SSBRI) can be reported by a UE based on the existing framework for CSI reporting as baseline. The predicted Top-K beams can be determined at least based on L1-RSRP(s) and a threshold  Proposal 19: For UE-sided model in BM-Case2, it should be considered to report predicted beams of multiple future time instances in one reporting instance  based on existing CSI-reportConfig as baseline  Proposal 20: For UE-sided model in BM-Case2, the time stamp of future time instance(s) can be implicitly indicated to NW.\ |
| Huawei/HiSi [33] | Proposal 19: For the CSI report for the inference of a UE-side AI/ML model, the predicted beam ID(s)/RSRP(s) and the measured beam ID(s)/RSRP(s) need to be differentiated, e.g., by introducing an indicator included in CSI-reportConfig.  Proposal 20: For BM-Case 2 with a UE-side model, the model output for N future time instances can be sent in one report.   * Overhead reduction techniques can be considered, e.g. model output compression with differential RSRP over temporal domain. * The time stamp of the reports can be derived implicitly from the order of the prediction instances.   Proposal 21: For BM-Case 2 with a UE-side model, investigate the supported resource types for the observation window:   * At least P/SP CSI-RS resources can be considered.   + Study whether/how to define the observation window.   + Study the CSI processing criteria to measure/process/store the observation instances. * Aperiodic CSI-RS resources may not be applicable due to long observation window.   Proposal 22: For BM-Case 2 with a UE-side model, for the reference time to determine the earliest time instance of the prediction window, consider Option 1 and Option 2 with higher priority.   * Option 1: Based on the time domain resource for the report. * Option 2: Based on the CSI reference resource corresponding to the report. * Time domain separation between two adjacent time instance(s) from the N future time instances are the same.   Proposal 23: For AI/ML model inference at the UE-side under BM-Case 1 and BM-Case 2, support to report the predicted beam IDs/RSRPs (i.e., Max value of K) 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 NW-side model.   Proposal 24: For AI/ML model inference at the UE-side under BM-Case 1 and BM-Case 2, to reduce the reporting overhead, consider to report a selected number of beams (i.e., Top-K value) determined by the UE based on output RSRP/probability threshold. |
| Xiaomi [34] | Proposal 4-1: Support to report the predicted L1-RSRP if the beam is not configured for corresponding measurement, and report the measured L1-RSRP if the beam is configured for corresponding measurement for UE-side model inference.  Proposal 4-7: If the measured beam information of the last history measurement time instance is reported, support to report the predicted beam information together for UE-side model inference in BM Case 2.  Proposal 4-8: Consider one absolute L1-RSRP for each predicted time instance or one absolute L1-RSRP for all predicted time instances in one beam report for UE-side model inference in BM-case 2. |
| Kyocera [35] | Proposal 7: For the UE-side AI/ML model, regarding the content of the report inference results, support using options 1 through 4. Further study the benefits and gains of adopting options 3 and 4, considering the effects of quantization methods on system throughput and considering the additional overhead incurred.  Proposal 8: For a UE side AI/ML model, for the FFS on beam information, support using CRI/SSBRI as a starting point.  Proposal 9: For a UE-side AI/ML model, the definition of the predicted top-K beams is related to the definition of the data set labels and the method of label collection. If the model output is specified (e.g., beam information such as CRI/SSBRI), it may be unnecessary to define the top-K predicted beams, and how they are generated during training, explicitly, and this could be left to implementation. |
| LGE [36] | Proposal #9: For supported Option 1 and Option 2, support K=4 for the max value.   * Considering the case that predicted beam is in Set A but not in Set B, reported beam information can be based on the relation/association between Set A beams and Set B beams   Proposal #10: For predicted RSRP report, confidence/probability information may be helpful for NW to decide whether/how to use the reported RSRP. Further study whether the information is per model/functionality, per report, per time instance, or per report parameter.  Proposal #11: Support to report inference results of N(N>=1) future time instance(s) in one report.   * Maximum value of N can be more than 1 * Maximum value of N can be reported by UE capability, and M(M<=N) value for inference results reporting can be configured by gNB * Further consider enhancement on RSRP quantization for UCI overhead reduction   Proposal #12: For temporal DL Tx beam prediction, information on time-variation of RSRP can also be included in the report.  Proposal #13: For temporal DL Tx beam prediction 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   Proposal #14: Support Option A or Option C (new) for the RSRP of predicted Top K beam(s) in the report of inference results.   * Option A: Predicted RSRP * Option C: Not specify whether to report predicted RSRP or measured L1-RSRP when both RSRPs are available at UE side, i.e., leave this case up to UE implementation   Proposal #15: For UE-sided model inference, consider reporting of performance monitoring related information such as actual Set A measurement or performance monitoring output/result together with predicted beam information on Set A, where the performance monitoring related information may be reported with longer periodicity. |
| NEC [38] | Proposal 10: If the AI/ML based CSI report is introduced, define the priority rule at least in following cases:  − Non-AI/ML based CSI report vs. AI/ML based CSI report  − AI/ML based CSI report vs. AI/ML based CSI report  Proposal 18: For report content of inference results for UE-sided model for BM-Case 1, for the RSRP of predicted Top K beam(s) in the report of inference results, Option B should be supported, i.e., Predicted RSRP, if the beam is not configured for corresponding measurement, and measured L1-RSRP if the beam is configured for corresponding measurement.  Proposal 19: For predicted RSRP, the Tx power is assumed based on the configured powerControlOffsetSS of the resource corresponding to the predicted beam if Set A resources are configured and the Tx power is assumed based on setting powerControlOffsetSS to 0 if Set A resources are not configured. |
| MTK [39] | Proposal 32: For BM-Case1 and BM-Case2 with a UE-side AI/ML model, study how to facilitate UE to report various number of Top-K beams in one beam report (K ≤ nrofReportedRS) as AI/ML model output. |

### Issue #6.1(on hold): Inference result report for BM-Case 2

### Issue #6.2: Inference result report for BM-Case 1

Summary from contributions:

For UE-sided model, at least for BM-Case1, for content in the report of inference results,

* Opt 3: Beam information on predicted Top K beam(s) among a set of beams and probability information of predicted Top K beam(s) among a set of beams
  + FFS on the quantization method of probability information
  + Probability information is the probability of the beam to be the Top 1 or Top K beam
    - Supported by: ~~Futurewei~~, GOOGLE, Ericsson, OPPO, Nokia, Sony,Samsung, Kyocera
* Opt 4: Beam information on predicted Top K beam(s) among a set of beams, RSRP of predicted Top K beam(s) among a set of beams, and confidence information of the RSRP
  + FFS on definition of reported RSRP
  + FFS on the definition and quantization method of confidence information
    - Supported by: ~~Futurewei~~. GOOGLE, Ericsson, Kyocera,LGE

##### (FL0) Proposal 6.2-1:

For UE-sided model, at least for BM-Case1, for content in the report of inference results, for Opt 1 (only beam information of predicted Top K beam(s)) further study the method to convey the ranking information of the predicted Top K beams for K > 1.

|  |  |
| --- | --- |
| Supported companies | FL |
| Companies | View |
| FL | This proposal makes sense. |
| HW/HiSi | Not sure if this proposal is needed for all cases.  If RSRP and beam indices are reported, the ranking can be directly seen from RSRPs. If only beam indices are reported, the ranking is still naturally conveyed by the index order. |
| TCL | If is are a beam sweeping among the predicted K beams to confirm the best one, then the ranking information is not required. |
| Xiaomi | ok |
| ZTE | Support |
| InterDigital | Do not support. The proposal is not needed. |
| Ericsson | Probably not needed. This would anyway be studied when the reporting format is discussed. One solution is that UE includes the probability of a beam being the top-1 beam. |
| CATT | OK with the proposal. It is natural to convey the ranking information by the order of reported index order. We can make a further step as the following:  For UE-sided model, at least for BM-Case1, for content in the report of inference results, for Opt 1 (only beam information of predicted Top K beam(s)), the ranking information of the predicted Top K beams for K > 1 can be conveyed via the order of reported index. |
| Nokia | Need some discussion to understand this. |
| Panasonic | We share similar view as HW/HiSi. |
| MediaTek | We are OK with this proposal. |
| CMCC | Fine to discuss. The ranking of the predicted Top K beams can be in a descending order of predicted RSRP so that gNB can know which is predicted top1 beam. |
| Sharp | Support. |
| Futurewei | Ok with the direction of the proposal. By the way, our views on Opt 3 and Opt 4 were incorrectly captured above. We are NOT supportive of Opt 3 and Opt 4. |
| NEC | support |
| Lenovo | Seems not needed. |

##### (FL1) Proposal 6.2-1a:

For UE-sided model, at least for BM-Case1, for Opt 1 (only beam information of predicted Top K beam(s)) the ranking information of the predicted Top K beams is conveyed by the ordering in the report, e.g., the first beam information is the one with the largest predicted RSRP or with largest probability of the beam to be the Top 1 or Top K beam

|  |  |
| --- | --- |
| Companies | View |
| FL | The original proposal is not clear enough. Hope this one can be more clear. |
| MediaTek | Ok with the direction but the example is too specific, we suggest the following wording:  For UE-sided model, at least for BM-Case1, for Opt 1 (only beam information of predicted Top K beam(s)) the ranking information of the predicted Top K beams is conveyed by the ordering in the report, e.g., the beam information in the report is ordered/ranked based on the beam’s predicted RSRP or probability. ~~the first beam information is the one with the largest predicted RSRP or with largest probability of the beam to be the Top 1 or Top K beam~~ |
| New H3C | OK in general |

##### (FL0) Proposal 6.2-2:

For UE-sided model, for the quantization of a predicted RSRP value at least for the report of reference results, support

* Support differential RSRP reporting with legacy quantization step and range for L1-RSRP reporting
  + For BM-Case 1, support differential RSRP report among multiple beams
  + For BM-Case 2, support differential RSRP report among multiple beams in one time instance and over multiple time instances

|  |  |
| --- | --- |
| Supported companies | FL, NTT DOCOMO |
| Companies | View |
| FL | For BM-Case 1, or within one time instance.  And all proposals support differential RSRP over multiple time instances. I want to give a try as well |
| OPPO | Support the FL proposal. |
| HW/HiSi | Support |
| Fujitsu | Generally fine with FL proposal. |
| New H3C | OK in general |
| TCL | Agree |
| ETRI | Support this proposal. |
| Xiaomi | For BM case 2, does ‘support differential RSRP report among multiple beams over multiple time instances’ mean support only one absolute L1-RSRP among multiple time instance? If yes, we support this proposal. |
| ZTE | For the main bullet, there's a typo where 'reference results' should be 'inference results'. Additionally, since it's still not decided yet whether the reported RSRP should be the predicted RSRP or the measured RSRP for beams within Set B, we suggest not to use 'predicted RSRP' in the main bullet for now.  For the last sub sub-bullet, there are still some different understandings for differential RSRP report over multiple time instances, and thus more details are to be discussed. For example, differential RSRP values of all time instances are reported with a reference to the largest RSRP values over all the multiple time instances. Or, differential RSRP values of each time instance are reported with a reference to the largest RSRP value of the same time instance as legacy, and those largest RSRP values of each time instance are reported referring to the largest RSRP value over all the multiple time instances.  Therefore, we suggest the following revisions.  For UE-sided model, for the quantization of a ~~predicted~~ RSRP value at least for the report of ~~reference~~ inference results, support   * Support differential RSRP reporting with legacy quantization step and range for L1-RSRP reporting   + For BM-Case 1, support differential RSRP report among multiple beams   + For BM-Case 2, support differential RSRP report among multiple beams ~~in one time instance and~~ over multiple time instances     - FFS details |
| Ericsson | Support |
| LG | Support in principle.  Question for BM-Case 2: do we support indication of beam index for the largest value of predicted RSRP per time instance or over multiple time instances? |
| CATT | Support the FL proposal. |
| vivo | Fine with the proposal |
| Fraunhofer | Support. |
| Nokia | OK |
| Panasonic | Support. |
| MediaTek | We have same question as Xiaomi, it is unclear to us whether the 2nd bullet means (1) Each past time instance in the report has one highest absolute L1-RSRP and differential L1-RSRP is applied to the rest of the beams in the same time instance, or (2) Differential L1-RSRP is defined as the L1-RSRP difference between the first time instance to each time instance for a same beam. |
| Spreadtrum | Support |
| CMCC | Support. |
| Kyocera | Support the FL proposal |
| Sharp | Support. |
| Futurewei | Support. |
| Lenovo | Support |
| Apple | Xiaomi’s clarification question is useful. We support modifying FL’s proposal accordingly |

##### (FL1) Proposal 6.2-2a:

For UE-sided model, for the quantization of a ~~predicted~~ RSRP value at least for the report of inference results, support

* Support differential RSRP reporting with legacy quantization step and range for L1-RSRP reporting
  + For BM-Case 1, support differential RSRP report among multiple beams
  + For BM-Case 2, support differential RSRP report among multiple beams ~~in one time instance and~~ over multiple time instances
    - FFS details

|  |  |
| --- | --- |
| Companies | View |
| FL | The original proposal is not clear enough. Hope this one can be clearer. |
| New H3C | OK in general |

### Issue #6.3: Definition of beam information in the inference report

##### (FL0) Proposal 6.3:

Beam information in inference report for UE-sided model refers to Set A

* FFS on details considering the options of content and configuration of resource set for Set A
  + [Beam information for measurement report for NW-sided model can be a starting point when applicable]

|  |  |
| --- | --- |
| Supported companies | FL |
| Companies | View |
| FL | I don’t expect concerns on the main bullet |
| OPPO | Yes, no concern over the main bullet. |
| HW/HiSi | Ok with main bullet |
| Fujitsu | For UE-side model, we think the beam information should include the information of different CCs, e.g., CC ID, especially for the generalized model for multiple CCs. |
| New H3C | OK in general |
| TCL | Agree |
| Xiaomi | Fine with the main bullet. |
| ZTE | Prefer to delete the last sub sub-bullet. There are some essential differences between the inference report for UE-sided model and measurement report for NW-sided model, including the report quantity and the number of beam to be reported. Therefore, the beam information for the two cases can be treated separately. For example, for the inference report for UE-sided model, it can be generally assumed that the number of beam to be reported is not greater than 4 as legacy, and then the beam information can be conveyed by CRI or SSBRI. However, for the measurement report for NW-sided model, the number of beam to be reported is generally far more than 4, and then some enhanced format for beam information (e.g., bitmap) is more preferable from reporting overhead perspective. |
| Ericsson | The proposal is unclear, is the intention that we should make such conclusion? |
| LG | What options do we have? |
| CATT | OK with the main bullet. Whether/how to configure Set A is still FFS, the bullet in the bracket can be removed. |
| Fraunhofer | OK |
| Nokia | The first bullet was agreed in an earlier meeting. |
| CMCC | Ok. |
| Futurewei | Fine with the main bullet. |
| Lenovo | We are fine with the main bullet. |
| Apple | The proposal is unclear. |

# Beam indication

### Summary from the contributions

|  |  |
| --- | --- |
|  |  |
| Spreadtrum [2] | Proposal 7: For BM-Case2, TCI indication framework should be reused by gNB, e.g., beams from multiple time instance can be indicated to UE by multiple beam indications respectively.  Proposal 8: For BM-Case 1 and BM-Case 2, support UE to report the measurement results of up to 16 beams in one reporting instance. |
| GOOGLE [3] | Proposal 7: Since the activated/indicated TCI based on SD beam prediction is usually an unknown TCI state, to reduce the latency for TCI activation/indication based on SD beam prediction, support the NW to trigger aperiodic CSI-RS resources QCLed with the SSB/CSI-RS configured as the QCL source in the TCI state.   * UE measures time/frequency offset and Rx beam based on the aperiodic CSI-RS resources * UE can also measure the pathloss based on the aperiodic CSI-RS resources   Proposal 19: To differentiate the TCI state for legacy beam indication and TCI state for beam prediction, support to configure separate TCI state pools for legacy beam indication and TCI state for beam prediction.  Proposal 20: Support to configure the action delay for the TCI state for beam prediction.  Proposal 21: For temporal 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 22: Support UE feedback before the beam action time for performance validation for predicted beam in addition to the ACK/NACK for the TCI update signaling for temporal beam prediction. |
| CMCC[5] | Proposal 13: Top K beam sweeping procedure can be introduced and is configurable by gNB.  Proposal 14: The indication of Top K beam set with low signaling overhead needs further discussion.  Proposal 20: The RS associated with TCI indication should be measured at least once before TCI application. TCI indication associated without RS in set A is not supported. |
| Intel [6] | Proposal 19: For BM-Case 1 and 2, RAN1 should consider beam indication of predicted beams which have TCI states that are not part of the set of MAC-CE activated TCI states. |
| ZTE [7] | Proposal 25: For BM-Case2 (both UE-sided and NW-sided model), support to extend the Rel-17 TCI state activation/indication signalling methods to activate/indicate N TCI states which are corresponding to N future time instances.  Proposal 26: Study enhanced QCL indication method for aperiodic RS resources for sweeping over the predicted Top-K beams. |
| Ericsson [8] | Proposal 7 For UE/NW-sided models, further study how Top-K beam measurements (P2 sweep) can be introduced in the existing CSI framework. With the following aspects as a starting point   * How NW indicates which beams in set A that are part of the Top-K measurements * How NW can configure a dynamic number K resources for measurements |
| OPPO [10] | Proposal 1: For BM-Case2 with NW-side model, enhance unified TCI framework to facilitate beam indication for multiple future time instances.  Proposal 12: For BM-Case2 with UE-side model, support to indicate multiple unified TCI states for up to F future time instances with one-shot beam indication. |
| Fujitsu [11] | * For BM Case-2 with UE side model, RAN1 to discuss beam indication enhancement, for example, TCI states of multiple time instances could be indicated via one DCI.   Proposal 29:   * For BM Case-2 with NW-side model, similar beam indication enhancement as BM Case-2 with UE side model could be considered. |
| CATT [12] | Proposal 12: For beam indication of BM-Case2, when studying TCI state indication of multiple future time instances using single indication signaling, the benefit, necessity, and TCI indication overwriting scheme should be considered. |
| TCL [13] | Proposal 6: The following enhancement on the TCI framework should be considered to support the AI/ML BM.   * Additional TCI state ID dedicated for AI/ML BP should be introduced. * New QCL types is indicated in TCI state to associate the RS sets corresponding to Set A and Set B beams. * At least for BM-Case2, timing information related to different predicted beams should be configured to the UE using RRC signaling, e.g., included in the TCI state information. |
| InterDigital [16] | **Proposal 4**: A RS resource, which is not transmitted, for Set A (i.e., unmeasured beams in Set A not in Set B), is configured with a TCI state including a logical beam ID as a QCL Type-D reference RS.  **Proposal 5**: To indicate Set A beams (i.e., unmeasured beams in Set A not in Set B), support indication of Set A beams based on the following options:   * Option 1: a TCI state using a logical beam ID as a QCL Type-D reference RS. * Option 2: a TCI state using a RS resource, which is configured with a logical beam ID as a QCL Type-D reference RS, as a QCL Type-D reference RS of the TCI state.   **Proposal 6**: Support a procedure for the UE to obtain QCL-parameters for an unmeasured Set A beam by using neighboring beams of the unmeasured Set A beam.  **Observation 8:** Enhanced beam indication mechanism is needed to enable future beam indication based on prediction of AI/ML model in BM-Case 2.  **Proposal 7**: Support a beam indication mechanism with a beam pattern and corresponding TCI states required for the indicated beam pattern. |
| Panasonic [17] | Proposal 3: For NW-sided model, support to extend Rel. 17 TCI state activation signaling methods to activate TCI states of *K* predicted beams for *N* future time instances in BM-Case 2. The following 2 options can be considered:   * Option 1: The TCI states of *K* predicted beams for *N* future time instances are included in a combined set of TCI states together with that of legacy BM. * Option 2: The TCI states of *K* predicted beams for *N* future time instances are included in a separate set of TCI states, compared to that of legacy BM. |
| Nokia [19] | Proposal 4: For BM-Case1 with the UE-sided model, consider following enhancements/limitations/changes related to the applicability of the beam indication.   * The applicability of the TCI indication for a channel/signal may be depended on whether the TCI indication is associated with a measured RS resource or predicted RS resource. E.g., extend *followUnifiedTCI-State*. * The UE considers a TCI indication associated with a predicted RS resource as known TCI state.   + Check the feasibility of this with RAN4   Proposal 5: For BM-Case2 with the UE-sided model, extend the Rel-17 TCI state activation/indication signalling methods to activate/indicate N TCI states which are corresponding to future time N instances.   * FFS: maximum number for N * FFS: Time periods that each indicated TCI state is applicable.   Proposal 6: For BM-Case2 with the UE-sided model, consider enhancements/limitations/changes related to applicability of the beam indication.   * The applicability of the TCI indications for a channel/signal may be depended on whether the TCI indications are associated with measured RS resources or predicted RS resources. E.g., extend *followUnifiedTCI-State*. * The UE considers TCI indications associated with predicted RS resources as known TCI state.   + Check the feasibility of this with RAN4   Proposal 21: For BM-Case2 with the NW-sided model, consider extending the Rel-17 TCI state activation/indication signalling methods to activate/indicate N TCI states corresponding to future time N instances. |
| Samsung [21] | Proposal 6. Support single beam indication for multiple future time instances using unified TCI framework.  Proposal 19. For UE-side AI/ML model, support beam indication for Set A beams. |
| Sharp [26] | Proposal 14 For BM-Case 2, NW-sided model and UE-sided model, enhance unified TCI state framework to support beam indication of multiple future time instances. |
| Qualcomm [27] | Proposal 7 For UE-side beam prediction, regarding FFS on beam information on predicted Top-K beams, conclude that such information includes beam indices from Set A.   * FFS: how UE reports such beam indices, considering the fact that Set A beams may not be based on RS’s that are actually transmitted. |
| Fraunhofer HHI, Fraunhofer IIS [29] | Proposal 17: Study whether beam indication for multiple future time instances is required. |
| ITL [31] | Proposal 12: For beam/TCI indication, consider using Set B beams of which UE can measure and maintain it Rx beam for P-3, if the gNB directs a beam within Set A that is unknown to the UE as the TCI state.  Proposal 13: For beam/TCI indication of BM-Case2(NW side model), consider extending the existing TCI direction method to multiple beams with the associated timestamp information for future time N instances. |
| Huawei/HiSi[33] | Proposal 15: For indicating the second round measurement of the predicted Top-K beams which may vary over time, study the following alternatives:   * Alt.1: Reuse legacy aperiodic CSI indication of the selected CSI report with pre-configured resource set of the Top-K beams. * Alt.2: Dynamic indication of RS resources which constitute the resource set of the Top-K beams. * Alt.3: Dynamic indication of the QCL relationship for the pre-configured RS resources of the Top-K beams.   Observation 3: In legacy, switching delay is needed for a TCI state to switch from unknown state to known state which has a corresponding testing requirement to UE.  Observation 4: For UE-side model, regarding the TCI state of a predicted beam which is subject to unknown state:   * If the model is testable at RAN4, it is up to RAN4 to decide whether to introduce a new TCI state/timeline for the predicted beam along with UE requirement. * If the model is not testable at RAN4, there will be no performance requirement for the predicted beam. * From RAN1 perspective, regardless whether it is testable or not, it does not preclude the gNB to schedule the unknown state TCI predicted by UE.   Proposal 16: For UE-side model, the existing definition/timeline of TCI states can be reused from RAN1 perspective as it does not cause ambiguity between UE and gNB on scheduling the predicted beam.  Proposal 17: For BM-Case 2, do not support to extend the Rel-17 TCI state activation/indication signaling methods to activate/indicate N TCI states which are corresponding to N future time instances, because   * Potential benefit of overhead saving (if any) is insignificant. * Overhead saving cannot be achieved under the following typical cases:   + For Top-K>1, second round beam sweeping would be anyway needed before the future time instance.   + PDSCH subject to the future time instance is scheduled by the same DCI indicating the TCI.   + gNB may need to update/override the TCI state that is previously predicted before each corresponding future time instance. * Substantial impact on implementation complexity and RAN4 impact (e.g. an increased number of active TCI states). |
| Xiaomi [34] | Proposal 4-2: UE to report whether the best Rx beam is known or not for the reported Tx beam to gNB. And for the beam without information of the best Rx beam, legacy procedure for Rx beam sweeping can be used to find the best Rx beam first and no new spec impact. Proposal 4-9: Support following two TCI state indication mechanism for TCI state indication of more than one predicted time instance.   * Option 1: reuse legacy TCI state indication with multiple MAC CE or DCI and each MAC CE or DCI indicates TCI state of one time instance. * Option 2: enhance TCI state indication to indicate TCI state of more than one time instance and the application time gap between two adjacent TCI states can be configured semi-statically. |
| NEC [38] | Proposal 36: To enhance unified TCI state to indicate the predicted beam, support to configure RS in associated Set A as the QCL reference signal in the TCI state.  Proposal 37: For BM-Case 2, support to use one MAC CE or DCI to activate/indicate multiple (future) TCI states, and corresponding time period. |
| MTK [39] | Proposal 25: For AI/ML-based BM, at this stage, there is no further enhancement needed for beam indication based on unified TCI state framework. |
| Meta [41] | Proposal 14: Consider enhancements to the unified TCI framework for indication of predicted beams which are not part of activated TCI states. |

### Issue #7.1: Whether and how to support beam indication for multiple further time instances

##### (FL0) Proposal 7.1:

For BM-Case2 (both UE-sided and NW-sided model), study on whether/how to extend the Rel-17 TCI state activation/indication signaling methods to activate/indicate N [joint] TCI states which are corresponding to N future time instances

* FFS: maximum number for N
* FFS: Time periods that each indicated TCI state is applicable.

|  |  |
| --- | --- |
| Supported companies | FL |
| Companies | View |
| FL | Only one companies expressed “NO” explicitly.  But the proposal is for study whether and how. |
| NTT DOCOMO | Support the proposal. |
| OPPO | Fine with this direction.  One minor comment on activated/indicated TCI state for N future time instances: for DL Tx beam prediction, it would be applicable to use joint or DL TCI states in our view. If that’s the case, could we suggest to add DL along with joint in the bracket?  to activate/indicate N [joint/DL] TCI states which are corresponding to N future time instances |
| HW/HiSi | Not support this proposal.  From our observation, more than just one company do not support this proposal. In addition to HW/HiSi, there are also for example MTK*: “Proposal 25: For AI/ML-based BM, at this stage, there is no further enhancement needed for beam indication based on unified TCI state framework.”*and Spreadtrum: “*Proposal 8: For BM-Case 1 and BM-Case 2, support UE to report the measurement results of up to 16 beams in one reporting instance.*”  In our paper, we have analyzed and studied the aspects related to activation/indication of multiple future time instances and come to the conclusion that it should not be introduced for the following reasons:   * The sole potential benefit with this proposal would overhead saving of DL control signaling. But this overhead saving would be very small:   + As an example, assume that BM-Case 2 predicts the Top-K beams for 4 predicted future time instances (N=4), where each instance has a duration of 80ms. Then, rather than using legacy and to send one DCI prior to each future time instance, it is proposed to use one common DCI that would indicate all 4 future TCI states. For this example, 3 DCIs during a period of 4\*80ms=320ms could be saved. However, this is a very small DCI overhead saving considering a long prediction window and applicable only to limited cases, while the resulting specification effort and implementation complexity would be substantial. * The overhead saving would only occur in corner cases, if no PDSCH would be scheduled, or if there is infrequent DL traffic, since otherwise a DCI is anyway transmitted in which the TCI state can be indicated. * In the TCI mechanism, only activated TCI states can be indicated by DCI. But legacy only supports 8 active TCI states, which may be too little if multiple future instances should be indicated from the active TCI state list. But an increased number of active TCI states would severely impact the UE complexity and needs RAN4 efforts. In MIMO Rel-18, this issue was addressed for multi-TRP and could not be agreed for the same UE complexity reasons. * The gNB may not always want to keep all the future TCI states it has predicted. E.g., the TCI state indicated to a UE for a future time instance may not only depend on the predicted beam of this single UE but also depend on other UE(s) to be paired with it in that future time, which cannot be predicted. Especially when the prediction window is long, the gNB may anyway need to send additional DCI to override the previously predicted TCI state. The mechanism of overriding also brings potential spec impact. * If the model output is Top-K>1 beams for each future time instance, which is a general case from the performance point of view, a second round beam sweeping would be needed immediately before each predicted instance. This diminishes the usefulness of indicating multiple future time instances even further.   It is therefore our view that the associated cost is too high and outweighs the potential benefit to support the indication of multiple future TCI states in one DCI. |
| Fujitsu | Generally fine with the FL proposal.  Ok with suggestion from OPPO to add DL TCI state. |
| TCL | Agree. New TCI state should be introduced to support the AI-purposed configuration. |
| Xiaomi | Fine with OPPO’s update. |
| ZTE | Fine |
| Ericsson | Not support.  The only scenario where the above would make sense is\_   * Top-1 beam is predicted for N time instances * Top-1 beam RSRP is predicted for N time instance (no measurement is needed for CQI) * There is no need to update the Top-1 beam prediction in time frame 1…N   Even if this scenario should be supported, the overhead reduction is minimal. The current method for activating TCI states for data reception via DCI is sufficient.  However, it is not clear how to configure the UE with the correct TCI states for Top-K beam measurements (issue 7.3). |
| LG | Support. |
| CATT | Not support.  The benefit and necessity should be discussed firstly. We think the benefit of extension rel-17 TCI is limited in case of having PDSCH transmission since anyway gNB should send a DCI scheduling PDSCH. And there is also TCI flexibility issue and the issue of how to overwrite the TCI state before it applies. At a result, we do have concern on extension rel-17 TCI to multiple time instances. |
| Fraunhofer | We believe this is not required. |
| Nokia | Ok |
| Panasonic | Support. |
| MediaTek | We agree with HW. Somehow our views are not listed here. We have studied the pros and cons for this feature and concluded that this feature is not essential, and the corresponding benefit is not clear and very limited, with the following observations:  (a) the overhead saving would only occur if model output is Top-1. If the model output is Top-K>1 beams for each future time instance, a P2 beam sweeping would be needed at each future time instance.  (c) RAN1 may need to specify another mechanism to handle the case when NW want to change a pre-indicated beam, and a new recover mechanism when the pre-indicated beam fails.  (d) it may potentially limit the capability of NW for choosing a different Tx beam based on newly reported beam information from other UEs during the future N time instances |
| Spreadtrum | Not support. Using multiple indication, gNB can select a more appropriate beam based on real-time channel changes. For example, if gNB predicts Top 3 beams for two future time instance (e.g., t1 and t2), gNB can execute beam sweeping before t1 for the first Top 3 beams, and then execute beam sweeping before t2 for the second Top 3 beams. Otherwise, gNB can only select the beam for t1/t2 only based on the prediction result |
| CMCC | Not support. TCI state activation/indication for each future instance is enough. |
| Sharp | Support. However, before determining whether to enhance the TCI indication for BM Case 2, we think time granularity of beam indication for BM Case 2 should be determined first. If we have to indicate TCI state for each slot, it is hard to meet the time requirement considering that HARQ-ACK should be transmitted first to confirm the reception of beam activation command. Thus, if the time granularity is quite small for beam indication, the enhancement for the beam indication should be applied to BM Case2. |
| Futurewei | Not support. The legacy unified TCI framework works and it is unclear about the benefit of the proposal as commented by other company. |
| NEC | Support |
| Kyocera | Support |
| Lenovo | Not Support.  Firstly, the benefit is not clear for us because the TCI state can be indicated by a PDCCH without DL assignment when there is no PDSCH scheduling.  Secondly, if multiple TCI states are indicted by a DCI for multiple future time instances, the additional timeline for the different beam application should be specified and RAN4 may need to be involved as well.  Finally, do the indicated multiple TCI states for future time instances should be activated before indication? |
| Apple | We don’t support this proposal |

### Issue #7.2(on hold): TCI indication associated without RS (in Set A)

Proposal 4: For BM-Case1 with the UE-sided model, consider following enhancements/limitations/changes related to the applicability of the beam indication.

* The applicability of the TCI indication for a channel/signal may be depended on whether the TCI indication is associated with a measured RS resource or predicted RS resource. E.g., extend *followUnifiedTCI-State*.
* The UE considers a TCI indication associated with a predicted RS resource as known TCI state.
  + Check the feasibility of this with RAN4

Proposal 15 For the beam indication FFS, first discuss:

• How TCI states of set A beams can be configured during training and inference,

• How/whether consistency in TCI states from training to inference can be ensured.

• How to ensure valid TCI states for Top-K measurements

CMCC [5]:

Proposal 20: The RS associated with TCI indication should be measured at least once before TCI application. TCI indication associated without RS in set A is not supported.

Intel [6]

Proposal 19: For BM-Case 1 and 2, RAN1 should consider beam indication of predicted beams which have TCI states that are not part of the set of MAC-CE activated TCI states.

### Issue #7.3(on hold): whether any enhancement is needed for P2?

Ref from Ericsson:



Figure 6: The large number of possible Top-K configurations with AI/ML is complex for NW to configure using legacy CSI framework

CMCC:

Proposal 13: Top K beam sweeping procedure can be introduced and is configurable by gNB.

One problem to configure Top K beam sweeping is Top K beams often change after each inference. If periodic/semi-persistent CSI report is used for Top K beam sweeping, set A is associated with the CSI report. A DL signaling is needed to indicate Top K beam subset for measurement before each report, signaling overhead is large. If aperiodic CSI report is used for Top K beam sweeping, one trigger state can be associated with multiple CSI report and multiple resource sets. Since the maximum number of trigger states for aperiodic CSI reporting and maximum number of resource sets associated with one trigger state is limited, it is difficult to configure all combinations of Top K beams and also occupies too many trigger states and resource sets oriented for other purposes. The method to indicated all combinations of Top K beams for aperiodic CSI reporting with low signaling overhead needs further discussion.

Proposal 14: The indication of Top K beam set with low signaling overhead needs further discussion.

# Others (low priority)

### Issue #8.1: For UE sided model, AI/ML processing capability

* Huawei/HiSi [32]:
* Proposal 33: The legacy CPU mechanism can be reused for AI/ML-based CSI processing, i.e., low priority CSI is not updated if simultaneously required CPU exceeds overall supported CPU.
  + No need to separate the AI/ML processing from the CSI measurement for the CPU counting.
  + FFS whether the overall CPU should be shared or separately counted between legacy CSI reporting and AI/ML-based CSI reporting, and among AI/ML features/functionalities.
* Proposal 32: Considering different complexities for different models/functionalities, AI/ML-based UE processing requirement may be reported for per functionality basis, including:
  + CSI processing unit of the functionality.
  + CSI processing timeline of the functionality.
  + Memory storage of the functionality.
* Proposal 34: For CSI processing timeline, both Zref and Z'ref can be considered for AI/ML-based CSI processing.
* Proposal 36: Discuss the memory occupancy alignment mechanism to align the availability of memory storage for updating the new AI/ML-based CSI report.
* Vivo [9] further study whether to define AI process capability including re-use or modified the existing CSI computation time and CSI processing units.
* Lenovo [16] Consider to introduce AI process units for AI based operation. Study the mechanism on how to determine the reported beams for beam report with AI/ML inference if there is no available AI/ML model inference processing resource.
* MTK [34] For UE-sided model, consider how to adapt current beamReportTiming framework/definition to include the AI/ML’s model inference delay.
* Fraunhofer [29] For UE-sided models, for inference, study the UE reporting its inference time to the gNB.
* DoCoMo [32] Proposal 11: Enhancements of CSI processing units should be considered for beam prediction.
* OPPO [10] Proposal 25: For BM-Case1 and BM-Case2, consider the UE capability from the aspects of Set B measurement and Set A prediction (for UE-side model only).
* NVIDIA [37] For AI/ML based beam prediction in spatial/time domain, introduce specification support for UE capability signaling for AI/ML based beam prediction including model training, model inference and model monitoring. Proposal 8: For AI/ML based beam prediction in spatial/time domain, introduce specification support for conditions of a Feature/FG available for functionality.

### Issue #8.2: Whether/how to address Measurement error

* Ericsson [2] The number of samples and statistical metrics of the performance metrics needs to be addressed.
* Intel [6] RAN1 should further discuss if one-shot L1 measurements are used for set B beams or if averaging of L1 measurements over time is needed.
* GOOGLE [3] Support SSB/CSI-RS repetition to improve the measurement accuracy for SD beam prediction. Support SSB/CSI-RS repetition to improve the measurement accuracy for temporal beam prediction.
* GOOGLE [3] Support to report a confidence level indicator for L1-RSRP report to indicate the maximum L1-RSRP measurement error for each beam.
* OPPO [9] For temporal domain beam prediction, suggest to study and evaluate the beam dwelling time prediction.
* Fujitsu [11] Regarding training data collection, repetition of the reference signals could be considered to improve the measurement accuracy and the same UE Rx beam should be maintained during the measurement.
* DoCoMo [25] Discuss how to handle measurement sensitivity issue in the measurements of Set B/C.

### Issue #8.3: How to define “a condition” for UE sided model

Xiaomi [34]

Proposal 2-1: BM Case 1 and BM Case 2 can be considered as different conditions for different functionalities.

Proposal 2-2: For BM case 1, different association/mapping between beams within set B and beams within set A can be considered as different conditions for different functionalities.

* + Condition 1: Set A and Set B are different (Set B is NOT a subset of Set A)
  + Condition 2: Set B is a subset of Set A (Set A and Set B are not the same)

Proposal 2-3: For BM Case 2, different repeat window can be considered as different conditions for different functionality.

Proposal 2-4: For each repeat window in BM case 2, different association/mapping between beams within set B and beams within set A can be considered as different conditions for different functionalities.

* + Condition 1: Set A and Set B are different (Set B is NOT a subset of Set A)
  + Condition 2: Set B is a subset of Set A (Set A and Set B are not the same)
  + Condition 3: Set A and Set B are the same

Proposal 2-5: Define different ranges of number of beams in set B and/ or set A as different conditions for different functionalities.

Proposal 2-6: Different content of model output can be considered as different conditions for different functionalities.

Proposal 2-7: Different performance metric and performance monitoring type can be considered as different conditions for different functionalities.

NEC [22]

Proposal 2: In addition to information of beams in Set A and Set B, the timing information of Set B measurements and Set A prediction occurrences should be specified as conditions for BM-Case1 and BM-Case2.

MediaTek [34]

Proposal 22: For BM-Case1 and BM-Case2 with a UE-side AI/ML model, for the BM-specific conditions regarding “information regarding model inference”, consider at least 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)

Proposal 23: For BM-Case1 and BM-Case2 with a UE-side AI/ML model, for the BM-specific conditions regarding “performance monitoring”, consider at least for the following sub-conditions,

• conditions on performance metrics

• conditions on the detectable events

Ruijie network [36]

Proposal 1: For N, it is configurable by gNB subject to UE capability.

Proposal 2: For the maximum number of N, it is subject to UE capability.

Proposal 3: K\_n is the same for each time instance n (n=1,2,…,N), i.e., K\_n=K, and K is configurable by gNB subject to UE capability.

For the max total number of sum of K\_n over N time instance(s), where Top K\_n beams(s) for time instance n, it should be configurable by gNB subject to UE capability.

Proposal 4: The max total number of sum of K\_n over N time instance(s) is configurable by gNB subject to UE capability.

Huawei/HiSi [33]

Proposal 37: For UE capability report of the condition for UE-side model, discuss the report of supported/needed configurations, including at least:

* The number of the needed data samples for training/monitoring.
* The supported configurations of RS/CSI report for Set A and/or Set B for model training/monitoring/inference.
* The supported values of Top-K for inference.
* Target performance and robustness information of the functionality.

### Issue #8.4: NW-sided consistency

CMCC [5]

Proposal 6: Rx beam assumption for a measurement report can be up to gNB implementation. If quasi-optimal Rx beam assumption applies, an indication may be needed in CSI-ReportConfig to align the Rx beam assumption between gNB and UE.

Intel [6]

Proposal 10: For data collection for a network-side model, further discuss the benefits of the UE reporting assistance information e.g., Rx beam assumption for L1 measurement for the beams configured for measurement and reporting.

Proposal 11: RAN1 should further discuss if and how consistency regarding UE Rx beam assumption can be maintained for measurements for training data collection and for measurement of set B for inference.

Proposal 27: For network-side AI/ML model, UE Rx beam assumptions for measuring sets A/B during training data generation may be considered part of additional conditions.

Fujitsu

Proposal 8:

* Regarding training data collection, the same UE Rx beam should be applied to the measurements on the reference signals for model input data (Set B) and the measurements on the reference signals for ground truth data (Set A).

Kyocera [35]

Proposal 13: For a NW-side AI/ML model, to ensure consistency between training and inference, study the feasibility of the following mechanisms:

• The UE reports the receive beam index to the NW.

• To eliminate the UL reporting overhead, the best receive beam is assumed by the NW. The NW configures the QCL relationship, and the UE uses the receive beam identified for the QCL source for performing the measurements.

NEC [38]

Proposal 38: Discuss how to reduce the signaling overhead of UE performing measurements of predicted beams to determine the suitable Rx beam.

Apple [40]

Proposal 5-2: *timeRestrictionForChannelMeasurements* or a new IE (timeRestrictionForHistoricChannelMeasurements for example) can be set to a numerical value to ensure NW and UE have the same understanding regarding Tx beam and Rx beam usage.

### Issue #8.5: Whether to introduce a state indicating unable to measure for NW-Side model

GOOGLE[3]

Proposal 4: Support to introduce a beam subset configuration indicating a subset of beams that the UE should always report in an L1-RSRP report

* Support to define one L1-RSRP state to indicate the reported beam is in an invalid state (UE is unable to measure the L1-RSRP that can meet the L1-RSRP measurement requirement)

OPPO [10]

Proposal 4: For BM-Case1 and BM-Case2 with NW-side model, it is NOT necessary to specify UE-side additional condition on UE Rx beam.

Nokia [19]

* FFS: how to indicate assumption on Rx beams in the CSI report.
  + “best” or “Quasi-optimal” Rx beam should be selected by the UE and reflected in the measurement reports.

CATT

Proposal 7: For NW-sided model, it is beneficial to align the Rx information of the measurements between network and UE.

Proposal 8: For NW-sided model, UE should use fixed Rx beam for each measured beam in Set A/Set B, and the specific Rx beam used for measurement is up to UE.

DCM[25]

Proposal 19: If NW side beam prediction gets difficult due to UE side additional condition (e.g., UE Rx beam assumption), some enhancements should be introduced.

LGE [36]

Proposal #5: In order to indicate one beam in Set A not in Set B, support indicating multiple neighboring beams from Set B for helping UE to find its Rx beam for the Set A beam.

# 9 Offline proposals

##### Proposal A:

* UE may assume the *similar properties* of a DL Tx beam or beam set/list associated with the same associated ID
  + FFS: whether/how to define *similar properties* of a DL Tx beam or beam set/list

##### Proposal 2.1b:

For BM-Case1 and BM-Case2 with a UE-side AI/ML model, for Type 1 performance monitoring Option 1 (NW-side performance monitoring), L1 signalling is used to send the measurement results to NW for the calculation of performance metrics at NW from RAN 1 perspective.

* FFS on further enhancement if needed

##### Proposal 6.2-2a:

For UE-sided model, for the quantization of a ~~predicted~~ RSRP value at least for the report of inference results, support

* Support differential RSRP reporting with legacy quantization step and range for L1-RSRP reporting
  + For BM-Case 1, support differential RSRP report among multiple beams
  + For BM-Case 2, support differential RSRP report among multiple beams ~~in one time instance and~~ over multiple time instances
    - FFS details

##### Proposal 4.2

For UE-sided model at least for BM Case-1, for a CSI report is used for the configuration of inference results reporting, support:

* + Two resource sets can be configured for Set A and Set B separately
    - FFS whether support only one resource set for Set B
  + UE only perform measurement on the resource set for Set B (i.e., *resourcesForChannel*, *resourcesForChannelMeasurement*)

The beam information in the inference report refers to the resource set for Set A

##### Proposal 5.7:

For NW-sided model, for inference, at least for BM-Case 1, the content in a beam report in L1 signaling, the following options for beam information when M<the size of measurement resource are considered:

* Option 1: CRI/SSBRI of a measurement resource set
* FFS: Option 2: a bit map to indicate the reported beams of a measurement resource set and one beam index (i.e., CRI/SSBRI) for the largest measured value of L1-RSRP of a measurement resource set
  + Note: combination of option 1 and option 2 are not precluded and condition (if any) can be studied

Note: Purpose, such as above “For NW-sided model, for inference, at least for BM-Case 1”, will not be specified in RAN 1 specifications

##### Proposal B (fall back of Proposal A):

* The associated ID is used to ensure consistency between training and inference regarding NW-side additional conditions for inference at UE
* RAN 1 has no consensus on whether/how to define NW-side additional conditions or UE assumption with the same associated ID

# 9 Reference

1. R1-2405808 Discussion on specification support for AI/ML-based beam management FUTUREWEI
2. R1-2405899 Discussion on AIML for beam management Spreadtrum Communications
3. R1-2405950 AI/ML based Beam Management Google
4. R1-2405963 AI/ML for Beam Management Tejas Networks Limited
5. R1-2405975 Discussion on specification support for beam management CMCC
6. R1-2406014 Specification support for AI/ML for beam management Intel Corporation
7. R1-2406054 Discussion on AI/ML-based beam management ZTE Corporation, Sanechips
8. R1-2406141 AI/ML for beam management Ericsson
9. R1-2406172 Specification support for beam management vivo
10. R1-2406254 On specification for AI/ML-based beam management OPPO
11. R1-2406305 Discussion on specification support on AI/ML for beam management Fujitsu
12. R1-2406353 Specification support for AI/ML-based beam management CATT
13. R1-2406395 Discussion on AIML Beam Management TCL
14. R1-2406440 AI/ML specification support for beam management Lenovo
15. R1-2406463 Discussions on AI/ML for beam management Sony
16. R1-2406497 Discussion on AI/ML for beam management InterDigital, Inc.
17. R1-2406526 Discussion on specification support for beam management Panasonic
18. R1-2406571 Discussion on AI/ML based beam management Hyundai Motor Company
19. R1-2406586 AI/ML for Beam Management Nokia
20. R1-2406593 Discussions on specification support for beam management Ruijie Networks Co. Ltd
21. R1-2406637 Discussion for supporting AI/ML based beam management Samsung
22. R1-2406699 Discussion on specification support for AI/ML beam management Transsion Holdings
23. R1-2406718 Discussion on specification support for beam management ETRI
24. R1-2406894 Discussions on AI/ML for beam management CAICT
25. R1-2406920 Discussion on AI/ML for beam management NTT DOCOMO, INC.
26. R1-2406969 Discussion on specification support for beam management Sharp
27. R1-2407019 Specification support for AI-ML-based beam management Qualcomm Incorporated
28. R1-2407064 Discussions on Specification Support of AI/ML for Beam Management Indian Institute of Tech (M), IIT Kanpur
29. R1-2407109 Specification support for beam management Fraunhofer HHI, Fraunhofer IIS
30. R1-2407116 A Novel Model-ID Free Approach for Interoperability in AI/ML Beam Management Use Cases NTU
31. R1-2407120 Specification support for AI/ML beam management ITL
32. R1-2407142 Specification support for beam management KDDI Corporation
33. R1-2406977 Discussion on beam management for AI/ML Huawei, HiSilicon
34. R1-2406269 Specification support for beam management Xiaomi
35. R1-2405944 Specification Support for AI/ML for Beam Management Kyocera
36. R1-2406416 Discussions on AI/ML for beam management LG Electronics
37. R1-2406492 Specification support for AI-enabled beam management NVIDIA
38. R1-2406541 Discussion on specification support for beam management NEC
39. R1-2406765 Discussion on specification support for AIML-based beam management MediaTek Inc.
40. R1-2406826 Discussion on AI/ML beam management Apple
41. R1-2406884 Discussion on AI/ML based beam management KT Corp.
42. R1-2406888 AI/ML for Beam Management Meta Ireland

# 10 Previous agreements

## 10.1 Agreement in RAN 1 #116

Agreement

For NW-sided model, for inference, in a beam report initiated by network, based on one measurement resource set, support the report of more than 4 beam related information in L1 signaling

* Note: Purpose, such as above “For NW-sided model, for inference”, will not be specified in RAN 1 specifications
* FFS on the report content for beam related information
* FFS on max number of reported beam related information in one report

Agreement

For UE-sided model, at least for BM-Case1, for content in the report of inference results, support

* Opt 1: Beam information on predicted Top K beam(s) among a set of beams
* Opt 2: Beam information on predicted Top K beam(s) among a set of beams and RSRP of predicted Top K beam(s) among a set of beams
* At least K=1 and more, FFS on max value
* FFS on beam information
* FFS on the definition of predicted Top K beam(s)
* FFS on definition of reported RSRP when applicable
* FFS on other information in the report with potential down selection among the following options
* Opt 3: Beam information on predicted Top K beam(s) among a set of beams and probability information of predicted Top K beam(s) among a set of beams
  + FFS on the quantization method of probability information
  + Probability information is the probability of the beam to be the Top 1 or Top K beam
* Opt 4: Beam information on predicted Top K beam(s) among a set of beams, RSRP of predicted Top K beam(s) among a set of beams, and confidence information of the RSRP
  + FFS on definition of reported RSRP
  + FFS on the definition and quantization method of confidence information
* Other options are not precluded.

where the set of beams is Set A, i.e., the beams for UE prediction.

Agreement

* For NW-sided model and for UE-sided model, beam indication is based on unified TCI state framework
* FFS on whether/how potential enhancement is needed

Conclusion

For UE sided model at least for inference, for measurement, the configuration of Set B,

* take the current CSI framework as the starting point

## 10.2 Agreement in RAN 1 #116b

Agreement

For UE-side AI/ML model inference, for BM-Case2, support to report inference results of N(N>=1, FFS on N) future time instance(s) in one report

* wherein information of inference results of one time instance is as in one report for BM-Case 1
  + Note: overhead reduction is not precluded
* FFS on details

Agreement

For network-sided AI/ML model for BM-Case1 and BM-Case2,

* support using existing CSI framework for configuration of Set A as the starting point
* support using existing CSI framework for configuration of Set B as the starting point
* Note: Purpose, such as above “For NW-sided model, for BM-Case1 and BM-Case2” and “Set A” and “Set B”, will not be specified in RAN 1 specifications

Agreement

For report content of inference results for UE-sided model for BM-Case 1, for the RSRP ofpredicted Top K beam(s) in the report of inference results, when applicable, further study the following options:

* Option A: Predicted RSRP
* Option B: Predicted RSRP, if the beam is not configured for corresponding measurement, and measured L1-RSRP if the beam is configured for corresponding measurement
* Where the predicted RSRP is based on AI/ML output
* Note: Support both Option A and Option B is not precluded.

Working Assumption

For report content of inference results for UE-sided model for BM-Case 2, the RSRP ofpredicted beam(s) in the report of inference results, is the predicted RSRP, where the predicted RSRP is based on AI/ML output

Agreement

For UE-sided model at least for BM Case-1, *CSI-ReportConfig* is used for the configuration of inference results reporting

* FFS on the details in the *CSI-ReportConfig*, at least considering:
  + Alt 1: one *CSI-ResourceConfigId* is configured for Set B
    - FFS: how UE can determine the information about set A
  + Alt 2: one *CSI-ResourceConfigId* is configured for both Set A and Set B
    - FFS: How to configure resource set(s) for Set A and Set B in *CSI-ResourceConfig*
  + Alt 3: two *CSI-ResourceConfigId* s are configured for Set A and Set B separately
  + Alt 4: one *CSI-ResourceConfigId* is configured for Set B, Set A is configured using separate resource set(s) other than that represented by *CSI-ResourceConfigId*
    - FFS: how to configure/indicate separate resource set(s) for Set A
  + Note: separate *CSI-ReportConfig* for Set A and Set B are not precluded.
  + Note: Not perform measurement for Set A and only perform measurement for Set B subject to the *CSI-ReportConfig*
  + FFS on the association between Set A and Set B with or without additional IE
  + Other necessary configuration are not precluded.

Agreement

Further study, for the consistency of NW-side additional condition across training and inference for UE-sided model for BM-Case 1 and BM Case 2, where the NW-side additional condition may at least impact UE assumption on beams of Set A/Set B:

* Opt1: Based on associated ID (Referring to AI 9.1.3.3)
  + FFS on what can be assumed by UE with the same associated ID across training and inference
  + FFS on how associated ID is introduced, e.g., within CSI framework, or outside of CSI framework
* Opt 2: Performance monitoring based
  + FFS details
* Other options are not precluded.

## 10.3 Agreement in RAN 1 #117

Agreement

For BM-Case1 and BM-Case2 with a UE-side AI/ML model:

* Support Type 1 performance monitoring, including the following two options:
  + Option 1 (NW-side performance monitoring):
    - UE sends a report to NW (for the calculation of performance metric at NW)
      * Measurement results from resource set for monitoring, e.g., L1-RSRP and/or RS index is supported as the content of the report
      * FFS on other contents
    - The report is at least configured/triggered by NW
    - Note: this may or may not have additional spec impact
  + Option 2 (UE-assisted performance monitoring):
    - UE calculates performance metric(s)
      * FFS how to report and what to report
  + FFS whether to trigger the report based on event(s) for Option 1 and/or Option 2
* FFS Type 2 performance monitoring

Agreement

At least for NW sided model, for the quantization of a reported L1-RSRP value at least for the report in L1 signaling, support

* Support differential L1-RSRP reporting with legacy quantization step and range
  + FFS: larger quantization step(s) than the already supported legacy quantization step for differential L1-RSRP and/or for absolute L1-RSRP
  + FFS: Smaller range(s) for differential L1-RSRP than the already supported legacy range

Agreement

Following Working Assumption is confirmed.

Working Assumption

For report content of inference results for UE-sided model for BM-Case 2, the RSRP ofpredicted beam(s) in the report of inference results, is the predicted RSRP, where the predicted RSRP is based on AI/ML output.

Agreement

For NW-sided model, for inference report, at least for BM-Case 1, the content in a beam report in L1 signaling, support

* L1-RSRPs and corresponding beam information of Top M beam(s) with largest M measured value(s) of L1-RSRP(s) of a measurement resource set, where M is configured by gNB
* If M = the size of the measurement resource set, the content is all L1-RSRPs and one beam index (i.e., CRI/SSBRI) for the largest measured value of L1-RSRP of a measurement resource set
* FFS: L1-RSRPs and corresponding beam information of up to M beams within X dB gap to the largest measured value of L1-RSRP, X and M are configured by gNB, and whether/how to report number of reported beams
* FFS on the maximum value of M (where M can be larger than 4) based on UE capability (M may or may not be different for different reporting contents)
* FFS on beam information
* Note: Purpose, such as above “For NW-sided model, for inference report, at least for BM-Case 1”, will not be specified in RAN 1 specifications