**3GPP TSG RAN WG1#117 R1-240xxxx**

**Fukuoka City, Fukuoka, Japan, May 20th – 24th, 2024**

**Agenda item:** 9.1.1

**Source:** Samsung (Moderator)

**Title:** FL summary #1 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 #116 on AI/ML for beam management.

## 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 | |

## 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 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   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 |
| Report for training | FFS | NA |
| 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 | Metrics and Procedures |

## 2 Performance monitoring

### 2.1 Metrics

Summary of the Tdoc:

|  |  |
| --- | --- |
| Company | Proposal/analysis/observation |
| Ericsson [2] | Proposal 17 For UE-sided model performance monitoring, further study the applicability, details and need for the following monitoring procedures,  a. Inference-based monitoring, e.g.  i. L1-RSRP error estimated from Top-K measurement or DMRS (data reception)  ii. Top-1 beam accuracy estimated from Top-K beam measurements (during P2)  iii. Probabilities in Top-1 beam predictions (from model inference)  b. Dedicated BM monitoring procedure (e.g. reuse training data collection)  i. NW transmits set A, or a subset of set A beams.  c. No dedicated BM monitoring procedure, but UE provides link quality related KPIs  Proposal 18 For UE-sided model performance metric report, support both per-sample based and aggregated reporting (e.g. 10th, 50th, 90th percentile of L1-RSRP error).  • FFS: Number of samples needed for the aggregated report |
| Huawei/HiSi[3] | Proposal 31: Consider the following options for further study on their applicability for performance monitoring for UE-sided model:   * Type 1, Option 1, UE report the following for NW to calculate the metrics:   + Alt1-1: Predicted Top 1 or Top K beams, and ground truth of the target Set A resources (and at the target time instance(s) for BMcase-2)   + Alt 2-1: Measured L1-RSRP of the configured resource(s)   + Note: Alt 1-1 and Alt 2-1 may or may not have difference from the measurement report for NW-sided model   + ~~Alt 4-1: Measured L1-RSRP, and the predicted RSRP of the configured resource(s) according to beam(s) in the same target Set A resources~~   + ~~Alt 4-2: measured [L1-]RSRP of current and predicted RSRP of the predicted Top 1 beam~~ * Type 1, Option 2, UE calculate the metric(s) and report the metric(s) to NW, or reports the event determined based on the calculated metric, where the type of metrics include:   + ~~All above alternatives~~   + Beam index prediction accuracy information   + Measured RSRP gap information * ~~Type 1, Option 2, considering the following alternatives that may define an event~~    + ~~Alt 1-2, Alt 2-2, Alt 3-1, Alt 3-2, Alt 4-1, Alt 4-2.~~ * Type 2, define threshold according to some metric(s) for UE to make decision(s) of functionality ~~model~~ selection/activation/deactivation/switching/fallback operation, where the type of metrics include:   + Beam index prediction accuracy information   + Measured RSRP gap information |
| Intel [4] | Proposal 18: 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., throughput, L1-RSRP, L1-SINR, hypothetical BLER.  • Alt.4: The L1-RSRP difference evaluated by comparing measured RSRP and predicted RSRP. |
| Samsung [8] | **Proposal 20. For performance monitoring of AI/ML model, at least for BM-Case1, consider the following metrics for further study.**   * **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.** |
| Vivo [9] | **Proposal 40: Support Alt. 1 and Alt.4, i.e. Beam prediction accuracy related KPIs and the L1-RSRP difference evaluated by comparing measured RSRP and predicted RSRP, in 4 options of performance metrics for performance monitoring of AI-based DL Tx beam prediction. *=> FL: what is the definition of Alt 1?*** |
| CATT [12] | **Proposal 20: 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.** |
| China Telecom [13] | ***Proposal 6: For performance monitoring of BM-Case1 and BM-Case2, the following performance metrics can be supported：***   * ***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*** |
| CMCC [14] | **Proposal 18: Beam prediction accuracy related KPI can be used as the metric of model performance monitoring.**  **Proposal 19: 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 20: For UE-side monitoring and hybrid model monitoring of UE-side AI/ML model, monitoring mechanism of multiple benchmarks needs discussion to facilitate UE to perform model monitoring.** |
| Sony [15] | **L1-RSRP difference between the predicted and actual values should serve as the performance metric.** |
| Lenovo [16] | **Proposal 20: 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.** |
| Fujitsu [20] | * *Regarding UE-side monitoring (Type-2 performance 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.* |
| Xiaomi [21] | ***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) |
| ZTE [24] | ***Proposal 23: Support beam prediction accuracy related KPIs (i.e., Alt.1) as the primary performance metric for AI/ML performance monitoring.*** |
| CAICT [25] | ***Proposal 1: Support FL’s proposal as the starting point for performance monitoring metric. Accessibility should be considered for gene-aided beam information related metrics.*** |
| ETRI [27] | **Proposal 9: Support prediction accuracy and L1-RSRP difference as performance metrics.** |
| OPPO [29] | **Proposal 6: On the performance metric for NW-side model, at least support the beam prediction accuracy related KPIs (Alt.1).**  **Proposal 15: On the performance metric for model monitoring, strive to apply the same metric(s) for both UE-side model and NW-side model.** |
| Nokia [31] | **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.** |
| DoCoMo [32] | **Proposal 14: 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 15: Support reporting statistic values of observed performance metrics derived from multiple samples and/or events based on those statistic values.** |
| Sharp [33] | **Proposal 11:** Regarding the performance metrics of model monitoring, support Beam prediction accuracy related KPIs, e.g., Top-K/1 beam prediction accuracy.  **Proposal 12:** 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). |
| MediaTek [34] | ***Proposal 16: For performance monitoring, study the following metrics calculated at UE and/or gNB side:***   * ***Alt.1-1: Statistical results on beam prediction accuracy related KPIs, e.g., Top-K/1 beam prediction accuracy*** * ***Alt.1-2: Hypothetical on beam prediction accuracy related KPIs on a subset of Set A of beams, e.g., Top-K/1 beam prediction accuracy*** * ***(New) Alt. 1-3: beam prediction ranking/ordering accuracy, e.g., by comparing the ranking/ordering of the best beams derived from model output and the ranking of the best beams derived from measurement*** * ***Alt 2-1: Measured L1-RSRP of configured resource(s).*** * ***[Alt 2-2: Hypothetical L1-RSRP based on the configured resource(s)]*** * ***Alt 3-1: Probability information of the predicted beam to be the Top 1.*** * ***Alt 3-2: A confidence interval or prediction interval associated with predicted L1-RSRPs at a specific confidence level (e.g., 95%).*** * ***Alt 4-1: The ~~L1-~~RSRP difference between the measured L1-RSRP and predicted RSRP of a set of beams ~~according to beam(s) in the same target Set A resources~~, e.g.***   + ***The RSRP difference between the predicted Top 1 beam or Top K beam(s) in the configured full/subset of Set A resources***   + ***The RSRP difference between the genie-aided Top 1 beam or Top K beam(s) in the configured full/subset of Set A resources***   + ***The RSRP difference between all the beams in the configured subset of Set A resources*** * ***Alt 4-2: The ~~L1-~~RSRP difference between measured L1-RSRP of current beam and predicted RSRP of the predicted Top 1 beam*** |
| KT [35] | ***Proposal 6. Support L1-RSRP difference evaluated by comparing measured RSRP and predicted RSRP as a performance metric.***  ***Proposal 7. Discuss how to configure the predicted best beam(s) as a benchmark/reference for the performance comparison.*** |
| Qualcomm[37] | Proposal 8 **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:**   * **Top-K beam prediction accuracy with L1-RSRP margin**   + **Ratio of the number of performance monitoring instances for which the highest measured L1-RSRP of Top-K predicted beams is** **within a margin of measured L1-RSRP of best measured beam ID from Set A, to the total number of performance monitoring instances.** * **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 9 **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).** |
| CEWiT [39] | ***For performance monitoring, support Alt.1-1, Alt.2-1 and Alt.4-1, i.e., statistical results on beam prediction accuracy related KPIs, measured L1-RSRP of configured resource(s), and the L1-RSRP difference between the measured L1-RSRP and predicted RSRP according to beam(s) in the same target Set A resources for the metrics calculated at UE and/or gNB side.*** |
| Indian Institute of Technology Madras (IITM), IIT Kanpur[40] | ***Proposal 1: For performance monitoring of UE-side and/or gNB-side models, we recommend Alt. 1-1, Alt. 2-1, Alt. 3-2, Alt. 4-1.*** |
| KDDI [41] | ***Proposal 6: High priority support for Alt.1 and Alt.4 monitoring metric.***   * ***Alt.1: Beam prediction accuracy related KPIs*** * ***Alt.4: The L1-RSRP difference evaluated by comparing measured RSRP and predicted RSRP***   ***Proposal 7：We propose the following definition for Alt. 4:***  ***“The L1-RSRP difference between the predicted RSPR and measured RSRP for the same target Set A resources.”***  ***Proposal 8: A specific example of the monitoring procedure when using the metric in Alt. 4 is shown below:***   1. ***Sweep Set B beam.*** 2. ***Take the measurements associated with Set B beam as input and output the RSRP for the predicted Set A beam.*** 3. ***Report the relevant information for the upper Set A beam.*** 4. ***Sweep all or part of the reported Set A beam.*** 5. ***Calculate the difference between the measured and predicted Set A beam.*** |

##### **Summary of views on metrics**

For performance monitoring, study the following metrics calculated at UE and/or gNB side:

* Alt.1-1: Statistical results on beam prediction accuracy related KPIs, e.g., Top-K/1 beam prediction accuracy, beam prediction accuracy within 1 dB margin
  + *Supported by: Ericsson (*Top-1 beam accuracy estimated from Top-K beam measurements (during P2)*), Intel?, H3C? vivo? Xiaomi, ETRI, Nokia, CEWiT, IITM,KDDI?*
* Alt.1-2: Hypothetical on beam prediction accuracy related KPIs on a subset of Set A of beams, e.g., Top-K/1 beam prediction accuracy, based on configured resource(s)
  + *Supported by:MediaTek*
* Alt 1-3: 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.
  + *Supported by:Samsung*
* Alt. 1-4: beam prediction ranking/ordering accuracy, e.g., by comparing the ranking/ordering of the best beams derived from model output and the ranking of the best beams derived from measurement
  + *Supported by:MediaTek*
* Alt 2-1: Measured L1-RSRP of configured resource(s).
  + *Supported by: Futurewei, intel? CEWiT, IITM*
  + *Deprioritized:, Spreadtrum*
* ~~Alt 2-2: Hypothetical L1-RSRP based on the configured resource(s)~~
* Alt 2-3: 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.
  + *Supported by: Samsung, xiaomi?, Nokia*
* Alt 2-4: 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.
  + *Supported by: Nokia*
* Alt 2-5: considering L1-RSRP of monitoring RS resources, determining hypothetical BLER-like metrics based on the RS measurements, etc
  + *Supported by: Nokia*
* Alt 3-1: Probability information of the predicted beam to be the Top 1.
  + *Supported by: Ericsson*
  + *Deprioritized: Huawei/HiSi, Spreadtrum*
* Alt 3-2: A confidence interval or prediction interval associated with predicted L1-RSRPs at a specific confidence level (e.g., 95%).
  + *Supported by: Ericsson (For UE-sided model performance metric report, support both per-sample based and aggregated reporting (e.g. 10th, 50th, 90th percentile of L1-RSRP error).)* *IITM*
* Alt 3-3: The probability information of Top-1 beam of Set A is lower than a threshold value or not.
  + *Supported by: Samsung*
* Alt 4-1: The L1-RSRP difference between the measured [L1-]RSRP and predicted RSRP according of a Set of beam ~~to beam(s) in the same target Set A or Set B, e.g~~.
  + The RSRP difference between the predicted Top 1 beam or [average of] Top K beam(s)
  + The RSRP difference between the genie-aided Top 1 beam or [average of] Top K beam(s)
  + *Supported by: Ericsson (L1-RSRP error estimated from Top-K measurement or DMRS (data reception)), Intel? H3C?vivo?, CATT (of Set B), Sony?, MediaTek CEWiT* *IITM,KDDI*
  + *Deprioritized: Huawei/HiSi*
* Alt 4-2: The ~~L1-~~RSRP difference between measured [L1-]RSRP of current beam and predicted RSRP of the predicted Top 1 beam
  + *Supported by : Mediatek*
  + *Deprioritized: Huawei/HiSi*

### 2.2 For UE sided model

|  |
| --- |
| **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 |

|  |  |
| --- | --- |
| **Company** | **Comments/proposals/observations** |
| Futurewei [1] | * *For Rel-19 AI/ML-based BM, support UE reporting L1-RSRP and the corresponding CRI/SSBRI for performance monitoring of BM-Case1 and BM-Case2.* * *the UE can either report the performance metric(s) to NW or report an event to NW based on the performance metric(s)*   *On the other hand, reporting an event to NW based on the performance metric(s) should be avoided as it requires defining new event(s) in the spec, which most likely will take significant effort and time of the group to accomplish* |
| Huawei/HiSi[3] | Proposal 29: For UE-side model monitoring, consider all 3 options further and assess their potential spec impact:  • For Type 1 Option 1, UE reports the label and inference output.  • For Type 1 Option 2, UE reports calculated metric or event. The relevant spec impact for reporting calculated metric includes:  o The metric calculation approach, e.g., per sample report, or report of the statistical value.  o The type of metric, e.g., prediction accuracy, or RSRP gap to the benchmark.  • For Type 2, UE does not report CSI, but reports the monitoring decision (e.g., activation/deactivation/fallback) to NW.  • For Type 2, gNB may configure a threshold criterion to facilitate the UE to perform model monitoring.  Proposal 31: Consider the following options for further study on their applicability for performance monitoring for UE-sided model:   * Type 1, Option 1, UE report the following for NW to calculate the metrics:   + Alt1-1: Predicted Top 1 or Top K beams, and ground truth of the target Set A resources (and at the target time instance(s) for BMcase-2)   + Alt 2-1: Measured L1-RSRP of the configured resource(s)   + Note: Alt 1-1 and Alt 2-1 may or may not have difference from the measurement report for NW-sided model   + ~~Alt 4-1: Measured L1-RSRP, and the predicted RSRP of the configured resource(s) according to beam(s) in the same target Set A resources~~   + ~~Alt 4-2: measured [L1-]RSRP of current and predicted RSRP of the predicted Top 1 beam~~ * Type 1, Option 2, UE calculate the metric(s) and report the metric(s) to NW, or reports the event determined based on the calculated metric, where the type of metrics include:   + ~~All above alternatives~~   + Beam index prediction accuracy information   + Measured RSRP gap information * ~~Type 1, Option 2, considering the following alternatives that may define an event~~    + ~~Alt 1-2, Alt 2-2, Alt 3-1, Alt 3-2, Alt 4-1, Alt 4-2.~~ * Type 2, define threshold according to some metric(s) for UE to make decision(s) of functionality ~~model~~ selection/activation/deactivation/switching/fallback operation, where the type of metrics include:   + Beam index prediction accuracy information   Measured RSRP gap information |
| **Intel [4]** | Proposal 20: For UE-side AI/ML models, consider UE-event-triggered model monitoring.  Proposal 21: 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.  Proposal 22: For a UE-side AI/ML model, support indication of AI/ML model failure based on measurement on a set of configured resources and comparison with predicted L1 measurements from the AI/ML model. If the RSRP difference is below a threshold, UE can indicate model failure to the network.  Proposal 23: 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. *=>Motivation for multiple models:*  *the network may configure the UE to monitor the performance of multiple models and compare periodically such that the decision to switch or activate an alternate model may be taken by the network* |
| **Samsung [8]** | **Proposal 21. For UE-side AI/ML model, support both Type 1 performance monitoring and Type 2 performance monitoring.**  **Proposal 22. For Type 1 NW-side performance monitoring of UE-side AI/ML model, the existing CSI reporting mechanism and the CSI reporting mechanism for UE-side model inference are used as baseline.**   * **FFS: Enhancement/Extension to those mechanisms**   **Proposal 23. For Type 1 UE-side performance monitoring of UE-side AI/ML model, 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 24. For 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’.** |
| **CATT [12]** | **Proposal 23: For Type 1 performance monitoring of Option 1, whether the results are reported on per sample based or multiple samples based can be considered.**  **Proposal 24: For Type 1 performance monitoring of Option 2, consider the following aspects:**   * **the report contents, e.g., the statistic of the performance metric;** * **the specific events and uplink resources for event-based reporting.**   **Proposal 25: 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.** |
| **China Telecom [13]** | ***Proposal 7. For UE-side AI/ML model, support both Type 1 performance monitoring and Type 2 performance monitoring.*** |
| **CMCC [14]** | **Proposal 14: For UE-side AI/ML model, Type 1 performance monitoring is prioritized.**  **Proposal 15: 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 16: 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 specification impact of performance metric or event reporting, and configuration reporting are considered.** |
| **Lenovo [16]** | **Proposal 19: 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.** |
| **LGE [18]** | **Proposal #16: 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.**  **Proposal #17: Support event-triggered UE reporting for UE-sided AI/ML performance monitoring.** |
| **Panasonic [19]** | **Proposal 8: For performance monitoring for UE-sided model,**   * **For functionality-based LCM, support Alt. 1 (UE-side model monitoring) for model selection/activation/ deactivation/switching. NW-side model monitoring or hybrid monitoring can be used for functionality level monitoring in spite that what model is used within UE.** * **For model-ID-based LCM, support both Alts. 2 (NW-side model monitoring) and 3 (hybrid model monitoring)** |
| **Fujitsu [20]** | * *Regarding performance monitoring for BM Case-1 with UE side model, either the UE-side monitoring (Type-2 performance monitoring) or the hybrid monitoring* *(Option 2 of Type-1 performance monitoring) could be considered.* * *Regarding UE-side monitoring (Type-2 performance 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.* * *Regarding UE-side monitoring (Type-2 performance monitoring) for BM Case-1 with UE side model, RAN1 to further discuss the corresponding reference signal configuration.*   ***Proposal 18:***   * *Regarding hybrid monitoring (Option 2 of Type-1 performance 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 19:***   * *Regarding hybrid monitoring (Option 2 of Type-1 performance monitoring) for BM Case-1 with UE side model, RAN1 to further discuss the reference signal configuration and reporting enhancement.* |
| **Xiaomi [21]** | ***Proposal 5-3: For UE-side AI/ML model, Option 2 (UE-assisted performance monitoring) of Type 1 performance minoring should be supported at least. Option 1 (NW-side performance monitoring) of Type 1 performance minoring can be considered with enhancement on signaling overhead reduction.***  ***Proposal 5-4: If the performance metric is the L1-RSRP difference, not consider the beam configured for corresponding measurements during inference.***  ***Proposal 5-5: The number of the inference instances should be known by the UE for calculation of a statistic value of performance metric for performance monitoring.***  ***Proposal 5-6: For UE-side AI/ML model, NW-side initiated performance monitoring based on measurement/report configuration via RRC can be considered for Type 1 performance monitoring.***  ***Proposal 5-7: For UE-side AI/ML model, UE-side initiated performance monitoring based on SR and UL MAC CE can be considered for Type 2 performance monitoring.***  ***Proposal 5-8: 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 5-9: For UE-side AI/ML model with Type 2 performance monitoring, configure a threshold to assist UE to make the decision.*** |
| **NEC [22]** | ***Proposal 24: Support NW triggered performance monitoring. At least for UE-side model, support UE initiated performance monitoring.***  ***Proposal 25: Support UE to initiate performance monitoring of current AI/ML model if beam failure instance indication or out-of-sync indication is received.***  ***Proposal 26: Support UE to activate candidate models and to initiate performance monitoring of candidate models before model switching or model selection.***  ***Proposal 27: Study simultaneous performance monitoring for multiple candidate models.***  ***Proposal 28: 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 29: Monitoring based on data distribution should be supported.***  ***Proposal 31: It might be necessary to consider selecting a subset of Set A as the monitoring RS resource set.*** |
| **ZTE [24]** | ***Proposal 24: Support Type 1-2 performance monitoring (i.e., hybrid performance monitoring) as it achieves the best trade-off between reporting overhead and system performance assurance.***  ***Proposal 25: 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 26: Model/functionality failure detection should be based on monitoring results of several consecutive times within a predefined monitoring window.***  ***Proposal 27: UE reporting based on measurement of Set B can serve as an always-on fallback method to guarantee continuous services quality.*** |
| **CAICT [25]** | ***Proposal 2: FL’s proposal on UE-sided model monitoring procedure could be used as baseline. It is also fine to remove Alt 2-1 from Type 1 related options.*** |
| **ETRI [27]** | **Proposal 10: Support Type 1 performance monitoring method for performance monitoring of UE-sided model.**  **Proposal 11: Support the use of Monitoring IDs to differentiate monitoring processes for UE-sided models.** |
| **OPPO [[29]** | Assuming the metric is worse than a pre-defined threshold or trigger LCM-related event(s), it is up to NW to execute LCM procedure, e.g. switching to another AI/ML model or fallback to non-AI scheme.  **Proposal 16: For UE-side model monitoring of Type 1, prefer Option 2 (UE-assisted performance monitoring) over Option 1 (NW-side performance monitoring) from the aspects of UL overhead.**  **Proposal 17: For UE-side model, support Type 2 (LCM making decision made by UE) performance monitoring.**  **Proposal 18: For performance monitoring of UE-side model, study and specify (if necessary) the LCM-related event(s) which can be triggered by performance metric(s).** |
| Fraunhofer [30] | **Proposal 7: For UE-sided models, for monitoring, support a model monitoring configuration that defines measurement resources for all Set A beams.** |
| Nokia [31] | **Proposal 7: For BM-Case1 and BM-Case2, support NW-sided performance monitoring for beam prediction related CSI reporting, and further discuss the 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 preferred by the NW) within the legacy CSI reporting framework.** * **Case 2: Enhancements when using the same CSI reporting configuration for 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.**   **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 11: 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 15: There is no additional requirement for defining any new signalling for functionality LCM for beam prediction use case.** |
| DoCoMo [32] | **Proposal 13: Support at least type 1 performance monitoring for functionality/model ID activation/deactivation decisions.** |
| Sharp [33] | **Proposal 13:** For UE-side AI/ML model, both Type 1 performance monitoring with option 1 and option 2 and Type 2 performance monitoring are supported. |
| MediaTek [34] | ***Proposal 17: For performance monitoring of BM Case1 and BM Case2, for the set of beams configured for UE to measure and monitor, consider different sizes of subset of Set A with different configuring periodicities.***  ***Proposal 18: Considering the following applicability for further on performance monitoring for UE-sided model:***   * ***Type 1, Option 1, UE report the following for NW to calculate the metrics:***   + ***Alt1-1, Alt 1-2, Alt 1-3, Alt 2-1, Alt 3-1, Alt 3-2, ~~Alt 4-1, Alt 4-2~~***   + ***Note: Contents in the inference report can be reused to reduce reporting overhead, e.g., the predicted Top-1/K beam ID(s) for Alt1-1/1-2/1-3*** * ***Type 1, Option 2, UE calculate the metric(s) and report the metric(s) to NW:***   + ***All the alternatives*** * ***Type 1, Option 2, considering the following alternatives that may define an event:***    + ***All the alternatives*** * ***Type 2, define threshold according to some metric(s) for UE to make decision(s) of model selection/activation/ deactivation/switching/fallback operation:***   + ***All above alternatives***   ***Proposal 19: For BM-Case1 and BM-Case2 with a UE-side AI/ML model, when functionality-based LCM is applicable, identify for each AI/ML functionality whether it is feasible for UE to initiate LCM operation requests.***  ***Proposal 20: For BM-Case1 and BM-Case2 with a UE-side AI/ML model, model, to facilitate UE to detect a monitoring event for performance monitoring, considering NW signaling to UE the following aspects,***   * ***The performance metrics monitored for the event*** * ***The threshold of the performance metrics for determining the occurrence of the event*** * ***The number of samples of the occurrence instances required for determining the occurrence of the event, where the occurrence instances are the monitoring samples that the monitored metrics falls below a threshold*** * ***The number of monitoring samples required for determining the occurrence of the event*** * ***The frequency of each monitoring samples*** |
| KT[35] | ***Proposal 5. Support that UE calculates performance metric(s) for UE-sided model.*** |
| Qualcomm[37] | Proposal 8 **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:**   * **Top-K beam prediction accuracy with L1-RSRP margin**   + **Ratio of the number of performance monitoring instances for which the highest measured L1-RSRP of Top-K predicted beams is within a margin of measured L1-RSRP of best measured beam ID from Set A, to the total number of performance monitoring instances.** * **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 9 **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 10 **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 11 **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.** |
| CEWiT [39] | ***Proposal 2 For performance monitoring of NW-sided model, support Type 1 Option 1 and Type 1 Option 2, i.e., UE reports measurements to the NW for calculating the metrics and UE calculates the metrics and report either the metrics or an event to the NW.*** |
| IITM[40] | ***Proposal 2: For performance monitoring of UE-side and/or gNB-side models, we recommend Alt. 1-1, Alt. 2-1, Alt. 4-1.*** |
| KDDI [41] | ***Proposal 9: The following shows the processing after metric calculation on the UE side.*** ***There are two options for processing after metric calculation on the UE side:***   * ***Opt. 1: The UE reports the metric to the network, and the network issues management instructions to the UE based on the reported metric.*** * ***Opt. 2: The network pre-configures the UE with threshold values for operations such as fallback and switching. The UE then requests or reports an operation to the network according to the metric calculation results and threshold values.*** |

##### **Summary of views on performance monitoring of UE sided model**

Considering the following applicability for further on performance monitoring for UE-sided model:

Type 1, Option 1, UE report the following for NW to calculate the metrics:

* *Supported by: Ericsson*
* Alt 1-1: Predicted Top 1 or Top K beams, and ground truth of the target Set A resources (and at the target time instance(s) for BMcase-2)
  + *Supported by: Huawei/HiSi, Spreadtrum?*
* Alt 2-1: Measured L1-RSRP of the configured resource(s)
  + *Supported by: Futurewei? Huawei/HiSi, Spreadtrum?*
* Alt 3-1: Probability information of the predicted beam to be the Top 1.
  + *Supported by: Spreadtrum? => shall this belong to Type Option 2, that UE calculate it and report?*
* Alt 4-1: Measured L1-RSRP, and the predicted RSRP
  + Alt 4-1: of the configured resource(s) according to beam(s) in the same target Set A resources
  + *Supported by: Spreadtrum?*
* Alt 4-2: measured [L1-]RSRP of current and predicted RSRP of the predicted Top 1 beam
  + *Supported by:*

Type 1, Option 2, UE calculate the metric(s) and report the metric(s) to NW:

* All above alternatives
* Alt1-1: Statistical results on beam prediction accuracy related KPIs, e.g., Top-K/1 beam prediction accuracy, beam prediction accuracy within 1 dB margin
  + FFS on how to define/configure a window for UE to calculate the metric for Type 1 option1 if supported.
  + *Supported by: Fujitsu?xiaomi?*
* Alt 3-1
  + *Supported by: Spreadtrum?*
* Alt 4-1
  + *Supported by: Spreadtrum? FUjitus?*

Type 1, Option 2, considering the following alternatives that may define an event:

* Alt 1-2, Alt 2-2, Alt 3-1, Alt 3-2, Alt 4-1, Alt 4-2.
* *Supported by: Ericsson. Intel (L1 RSRP), Samsung (Alt 1-3, Alt 2-3, Alt 3-3), LGE, xiaomi Alt (1-1)*

Type 2, define threshold according to some metric(s) for UE to make decision(s) of model selection/activation/ deactivation/switching/fallback operation:

* All above alternatives
* *Supported by: Huawei (report monitoring decision, based on configuration from NW?), Intel (L1 RSRP, report AI/ML model failure )*

##### **Summary of views on performance monitoring-based consistency**

Other than the above, some analysis on performance monitoring-based method for consistency:

|  |  |
| --- | --- |
| Company | Proposals |
| Ericsson [2] | Proposal 3 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  From Ericsson’s paper [2]  Figure 2: Real-time monitoring procedure for addressing consistency.  Confined/Local environment (Region where a real-time monitoring  result is valid)  model inference  Monitoring  (estimate inference accuracy)  Activate model?  Time-between monitoring procedure |
| Huawei/HiSi [3] | *Proposal 15: For the consistency of NW-side additional condition across training and inference for UE-sided model, UE side performance monitoring can also be considered.* |
| Intel [4] | Proposal 26: For a UE sided AI/ML model, for consistency between training and inference, performance monitoring-based approaches should be deprioritized. |
| Interdigital [11] | *Proposal 16: Support both Opt.1 (based on an associated ID) and Opt.2 (performance monitoring based) for the consistency of NW-side additional conditions.* |
| CMCC [14] | Opt 2: Performance monitoring based |
| Fujistu [20] | ***Proposal 34:***  *Regarding the consistency across training and inference, the option based on performance monitoring is preferred.* |
| ZTE [24] | Observation 2: Performance monitoring-based approach can be useful for the alignment of NW-side additional conditions and communication environment changes with minor spec effort foreseen. |
| Nokia [31] | **Proposal 14: 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.** |
| DoCoMo [32] | **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.** |
| KT [35] | ***Proposal 3. Discuss the details for the performance monitoring-based as a mechanism to ensure consistency between training and inference for UE-side models.*** |

### 2.3 for NW sided model

|  |  |
| --- | --- |
| **TR 38.843**  For BM-Case1 and BM-Case2 with a NW-side AI/ML model  - Beam measurement and report for model monitoring  - UE reporting of beam measurement(s) based on a set of beams indicated by gNB.  - Signalling, e.g., RRC-based, L1-based.  - Note: This may or may not have specification impact.  - NW monitors the performance metric(s) and makes decision(s) of model selection/activation/ deactivation/switching/ fallback operation  - Note: Performance and UE complexity, power consumption should be considered. | |
| **Company** | Comments/views/proposals |
| **Huawei/HiSi [3]** | Monitoring performance metrics do not need to be specified for the NW-sided model. All potential discussions should be focused on the UE sided model.  Except for the data collection, there is no need to specify a monitoring procedure/metric for the NW-sided model. |
| **Intel [4]** | Proposal 19: For network-side AI/ML models, consider UE event triggered model monitoring with periodic or aperiodic measurement on set B of beams. |
| **Vivo [9]** | **Proposal 42: For model monitoring with NW-side model, support NW provides to UE the target (e.g., Top-1 or Top-k) beam information of a Set A.** |
| **Apple [10]** | **Proposal 3-3: L1 beam reporting for performance monitoring for NW-side model is supported.** |
| **CAT[12]** | **Proposal 21: 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 22: For performance monitoring of NW-sided model, UE can report multiple measurement results of benchmarks/reference in one reporting instance.** |
| **CMCC [14]** | **Proposal 13: NW-side monitoring of NW-side AI/ML model is supported.** |
| **Fujitsu [20]** | ***Proposal 21:***   * *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.* |
| **Xiaomi [21]** | ***Proposal 5-10: 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 5-11: 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.*** |
| **CAICT [25]** | ***Proposal 3: Alt 1-1 and Alt 2-1 should be considered for NW-sided model monitoring.*** |
| **OPPO [29]** | ***Proposal 7: For performance monitoring of NW-side model, configure Set B and/or Set A for UE to collect measurement results and report to NW.*** |
| **Nokia [31]** | **Proposal 18: 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.** |
| **MediaTek [34]** | ***Proposal 21: Considering the following reporting content and applicability to the monitoring metrics alternatives for performance monitoring for NW-sided model:***   * ***UE reports the L1-RSRP measurements of the configured full/subset of Set A of beams***   + ***Applicable to Alt 2-1, Alt 4-1, Alt 4-2*** * ***UE reports the Top-1/Top-K beam ID of the configured full/subset of Set A of beam***   + ***Applicable to Alt1-1, Alt1-2, Alt1-3*** |
| ***IITM [40]*** | *Proposal 4: For beam information in the report for NW-sided model, we recommend Opt 0, Opt 1 and Opt 4.* |
| *Panasonic [19]* | **Proposal 9: Group-based beam reporting is modified to support performance monitoring for NW-sided model.**  **Proposal 3: Group-based beam reporting is modified to support to report more than 4 beams in one report.** |

##### **Summary of views on performance monitoring of NW sided model**

Considering the following applicability for further on performance monitoring for NW-sided model:

* Alt1-1: Top 1 or Top K beam information of the target Set A resources (and at the target time instance(s) for BMcase-2) based on measurements
* Alt 2-1, Alt 4-1, Alt 4-2 Measured L1-RSRP of the configured resource(s)
  + Also can support Alt 1-1

### 2.4 1st Round discussion

#### 2.4.1 Performance monitoring for UE sided model

#### 2.4.1.1 Type 1 performance monitoring for UE sided model

Re-summary of the metrics and corresponding report.

* Option A: Report the measurement results (e.g. L1-RSRP and/or beam information) of one set of beams, configured by NW
  + Using existing CSI framework for configuration of the set of beams as the starting point
    - FFS on the details of measurement results: e,g,L1-RSRP and/or beam information
  + *Comments from FL:* 
    - *Assuming: that NW already obtain prediction results from inference.*
    - *NW may configure full or subset of Set A for performance monitoring*
    - *Mapping to Type 1 option 1, Alt 1-1 and Alt 2-1 for metrics. NW will calculate the metric, without specify the metric.*
* Option B: Report the beam prediction accuracy related information
  + FFS on metric of beam prediction accuracy related information: e.g., Top-K/1 beam prediction accuracy, beam prediction accuracy within 1 dB margin, whether the beam prediction accuracy is then a threshold or not.
    - FFS on how to quantize the metric
  + FFS on configuration to obtain the beam prediction accuracy related information, including
    - How to define a window to obtain the statistical results on from UE
    - How to configure resources (e.g., full or subset of Set A, and Set B) to obtain the ground truth (genie-aided Top 1/K beam and corresponding L1-RSRP) to calculate the related KPI
  + *Comments from FL:* 
    - *The way to obtain ground truth:*
      * *Based on measurements of Set A and Set B to obtain the genie-aided Top 1/K beam and corresponding L1-RSRP. Put results for Set B into AI, and obtain the Top 1/K beam prediction accuracy, etc.*
        + *For BM Case 2, Set B is from multiple time instances and Set A is from the targeted perdition time instance.*
        + *For BM Case 2, there is no time instances information. But the actual measurements are from time by time.*
    - *NW may? configure full or subset of Set A for performance monitoring*
    - *Mapping to Type 1 option 2, Alt 1-1 for metrics. UE will calculate the metric, metric needs to be specified.*
* Option C: Report the RSRP difference information between the measured and predicted
  + 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:* 
    - *The RSRP difference is not clear defined, as we discussed in SI phase,*
      * *For #1, Regardless the beams are predicted Top 1 or Top K or not, UE just report the predicted L1-RSRP and measured L1-RSRP. In this case, if NW doesn’t configure the expected beams, it may cause some issues.*
      * *For #2: this requires NW must configure the predicted Top 1 or Top K of beams?*
      * *For #3: don’t see much benefit comparing with Option A.*
    - *Mapping to Type 1 option 2, Alt 2-3/2-4/4-2 for metrics. UE will calculate the metric, metric needs to be specified.*
* Option D: Report probability information of the predicted beam of Top 1 or Top K beams
  + FFS on probability information and the quantization
  + #1: The probability information of predicted Top 1
  + #2: The probability information of each or sum of predicted Top Top-K beams.
  + #3: Beam information that the probability information comparing to a threshold.
  + *Comments from FL:* 
    - *If option 3 is supported, this can be used. On the other hand, not need to define the purpose of Option 3. Maybe we should support Opt 3, without specify the purpose.*
    - *Mapping to Type 1 option 2, Alt 3-1/3-3 for metrics. UE will calculate the metric, metric needs to be specified.*
    - *No need to configure additional resource.*
* Option E: Report confidence information of the predicted RSRP error
  + FFS on confidence information, e.g., e.g. 10th, 50th, 90th percentile of L1-RSRP error)
  + *Comments from FL:* 
    - *If option 4 is supported, this can be used. On the other hand, not need to define the purpose of Option 4. Maybe we should support Opt 4, without specify the purpose.*
    - *Mapping to Type 1 option 2, Alt 3-2for metrics. UE will calculate the metric, metric needs to be specified.*
    - *No need to configure additional resource.*
* Option F: Report hypothetical BLER-like metrics
  + FFS on details
  + *Comments from FL: Still no clear definition.*

#### Issue #1: content of the report

##### **(High)Proposal 2-1A:**

For UE-sided model, for both BM-Case 1 and BM-Case2, for Type 1 performance monitoring,

* Support (Option A) to report the measurement results (e.g. L1-RSRP and/or beam information) of one set of beams, configured by NW
  + Using existing CSI framework for configuration of the set of beams as the starting point
    - FFS on necessary change for the report of measurement results: e,g, beam information only, L1-RSRP only, beam information and L1-RSRP
    - Strive for a common design as the report for NW-sided model
  + Note: this may or may not have additional specification impact
* Support (Option D) to 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.
* FFS on other options, including:
  + Option B: Report the beam prediction accuracy related information
  + Option C: Report the RSRP difference information between the measured and predicted
  + Option E: Report confidence information of the predicted RSRP error
  + Option F: Report hypothetical BLER-like metrics
  + Option D: Report an event notification, FFS on events
* FFS on whether to define event(s) to trigger above report(s)
* Note: Purpose, such as above “for both BM-Case 1 and BM-Case2, for performance monitoring” and “Set A” and “Set B”, may or may not be specified in RAN 1 specifications

|  |  |
| --- | --- |
| Company | Comments |
| FL | Please indicate your supported options.  For option B, C, E, F, there are too many unclear part, e..g,   * For option B, one or two companies gave some examples on how this can work. * For option C, I don’t see much motivation, assumption predicted RSRP already been reported to NW. and it is very hard to define the resource/method to obtain grand truth. * For Option E, lack of supporting companies. Even the companies support opt 4 (5vs5)for inference results are less than opt 3. * For option F, not clear, Nokia pls give a self-contained proposal for other companies to support this.   I think by default, we can declare support of Option A. then further study on the necessary enhancement for the configuration and report in Option A.  For Option D =Opt 3 (13 vs 3+1) of inference results, some companies indicated that this can be supported for performance monitoring. Therefore, I treated Option D is widely supported.  To get your preferred option supported, please resolve concerns from FL and provide self-contained proposals |
| HW/HiSi | Support option A.  Option D requires more discussion:   * Shouldn’t we firstly agree whether the probability can be reported for inference and then come back to a discussion about monitoring? * But still, for Option D, even if probability reporting would be supported for inference, we should also discuss if it is meaningful for monitoring. We have our doubts. Assume that the model is not working well, then the predicted top beam(s) will not be the true best beams. But according to the model output, the probability that will be indicated for the predicted beams will still be good (otherwise these beam would not have been predicted by the model). Therefore, it seems questionable how much the probability of the predicted top beams can say about the model performance to find the best beams.   For the FFS options, we think that this can be discussed later, but initial feedback from our side would be that it may be mea   * Option B: Report the beam prediction accuracy related information   + => Question for understanding: what is meant with the related information for the beam prediction accuracy? * Option C: Report the RSRP difference information between the measured and predicted   + => Not supportive. Also, we doubt that it is required that the RSRP difference between measured and predicted becomes as small as possible. Even if they would differ, the model could still be working well. * Option E: Report confidence information of the predicted RSRP error   + => Not supportive. Unclear how to define confidence. Also this is dependent on the inference discussion. * Option F: Report hypothetical BLER-like metrics   + => Not clear what it means. * Option D: Report an event notification, FFS on events   => We would need to define details about event firstly and can then come back to this Option later. |
| TCL | We support **Option A**, but suggest to change the proposal as follows.   * Support (Option A) to report the measurement results (e.g. L1-RSRP ~~and/~~or beam information) of one set of beams, configured by NW   + Using existing CSI framework for configuration of the set of beams as the starting point     - FFS on necessary change for the report of measurement results: e,g, beam information only~~, L1-RSRP only, beam information and L1-RSRP~~     - ~~Strive for a common design as the report for NW-sided model~~   Firstly, suppose the NW obtained the predicted beam IDs, then reporting the measured Beam IDs using the same ranking order with the predicted ones enables the NW knows if the prediction is accuracy. Similarly, if the NW obtained the predicted RSRPs, then reporting the measured RSRP according to the reporting order of predicted ones also enables the NW to do the comparison without the beam ID.  Secondly, we do not think the second FFS is valid, since for NW-sided model, the measured RSRP together with the beam ID should be reported, that is different from the UE-sided model where only one of them is needed.  For other options, we suggest to remove some of them, and I think the last Option should be “G” not “D”.   * FFS on other options, including:   + Option B: Report the beam prediction accuracy related information   + ~~Option C: Report the RSRP difference information between the measured and predicted~~   + ~~Option E: Report confidence information of the predicted RSRP error~~   + ~~Option F: Report hypothetical BLER-like metrics~~   + ~~Option G: Report an event notification, FFS on events~~   More details of Option B should be provided to differentiated it from A and D.  Option C is no better than Option A, but loses a lot of flexibility. As Option C needs the UE to do the comparison, it implies the beams for prediction and measurement should be the same, and the NW cannot determine which one to measure until receiving the report of predicted beams.  It is hard to imagine how Option E works, is it directly the model output or calculated by the measured and/or the predicted RSRP? More details should be provided if companies want to keep this option.  Option F and Option G are metrics of long-term performance which affected by many aspects, which does not directly reflect the beam prediction accuracy. Unless there are crystal clear evidence shows that it is better than Option A, or it compensates for the fatal flaw (if any) of Option A if they are using together, then we may consider to support these two options. |
| vivo | |  | | --- | | 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) |   Firstly, at least the reporting performance metric should be agreed for UE side monitoring.  Secondly, in the case of options D/E, the determination of probability and confidence lacks a clear definition and the effectiveness and advantages of the approach are not evident compared to other options. Further, if these metrics are based on the output of an AI model, it is unclear how using the AI model's output to assess its performance is meaningful or effective.  Thirdly, the reporting overhead of option A is too large for us.  In this case, we propose to discuss performance metric first based on following TR or support B/C first for performance metric   |  | | --- | | For the performance monitoring of BM-Case1 and BM-Case2:  - Performance metric(s) with the following alternatives:  - Alt.1: Beam prediction accuracy related KPIs, e.g., Top-K/1 beam prediction accuracy  - Alt.2: Link quality related KPIs, e.g., throughput, L1-RSRP, L1-SINR, hypothetical BLER  - Alt.3: Performance metric based on input/output data distribution of AI/ML  - Alt.4: The L1-RSRP difference evaluated by comparing measured RSRP and predicted RSRP | |
| MediaTek | 1. We support Option B. From our understanding, Option B is for Type1 option2 monitoring while Option A is for Type 1 Option1 monitoring. Beam prediction accuracy related metrics need to be derived statistically, therefore, what is provided by FL can be a good start in high level,   * Option B: Report the beam prediction accuracy related metrics ~~information~~   + FFS on metric of beam prediction accuracy related information: e.g., Top-K/1 beam prediction accuracy, beam prediction accuracy within 1 dB margin, beam prediction ranking accuracy~~, whether the beam prediction accuracy is then a threshold or not~~ (<- this should be defined in the event).     - ~~FFS on how to quantize the metric~~   + FFS on configuration to obtain the beam prediction accuracy related information, e.g., how to define a monitoring window/frequency and how to configure resource   We support to remove FFS for Option B.  2. The support of Option D should wait until Opt3 of inference results are agreed (currently this Opt is an FFS). |
| QC | Here’s our views on the options:   * Option A: depending on the size of Set A (or subset of Set A) this option may lead to excessive reporting overhead, which may not be justified for the monitoring of UE-side models. It is true that such reporting would be needed for the monitoring of NW-side models, but the necessity of it for UE-side models is not clear. On the other hand, the issue that FL has brought up regarding the case in which the Top-K beams may not be within the RS set for performance monitoring (for Option C) also holds for this option. If NW can wait for UE-side inference results first, then it can do it for Option C as well. * Option D: it is not clear how this option would help with “monitoring”. How is “ground truth” incorporated in this option? * Option B: this beam prediction accuracy related information can be also defined with a margin, as we had in Rel-18 evaluations, e.g., Top-1 L1-RSRP accuracy with an L1-RSRP margin. * Option C: this is a natural option which shows that if we go with the predicted beam how much L1-RSRP degradation we will have compared to the actual best beam from Set A. The concern brought up by FL regarding “what if Top-K is not among the subset of beams from Set A” is applicable to all the other options as well. * Option E: Based on the example from FL, I think this is mainly related to statistics of the per-instance metrics (?). This could be applicable to Option C as well, and generally for every per-instance metric, we can compute some statistics and report.   Option D: support the direction. |
| CATT | We support Option A, Option B and Option C. Since Option A can be applied to Type-1 Option 1 performance monitoring, while Option B and Option C can be applied to Type-2 Option 2 performance monitoring. For option A, the UE requires to report all the L1-RSRP of resource set. And for option B and Option C, the UE only requires to report beam prediction accuracy or L1-RSRP difference, which is benefit to reporting overhead reduction.  Regarding to Set A, if the set of beams is Set A, Option A means amounts of RS transmissions and large reporting overhead. One potential method is to configure Set B beam, the UE reports ideal L1-RSRP of Set B and predicted L1-RSRP of Set B, and then network can calculate the difference between measured Set B and predicted Set B. This method can reduce RS transmission and measurement of UE for performance monitoring.  For Option B, the beam prediction accuracy related information can be Top-1/K beam prediction accuracy or Top-1 beam prediction accuracy within 1 dB margin. When gNB configures set A of beams for measurement, the UE can obtain ideal Top-1 beam, or the beams within 1dB margin of ideal Top-1 beam among set A beams. Based on measurement result and AI/ML inference result, the UE can calculate:   * the percentage of "the Top-1 ideal beam is Top-1 predicted beam or one of Top-K predicted beam" * the percentage of the Top-1 predicted beam "whose ideal L1-RSRP is within 1dB of the ideal L1-RSRP of the Top-1 genie-aided beam"   Option C is applied to regression model, the UE calculates L1-RSRP difference between measured L1-RSRP and predicted L1-RSRP of one beam. In our understanding, this beam should refer to any one beam in Set A. To avoid transmitting all beams of Set A, the L1-RSRP of difference of beam in Set B can be used as 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. **For Option C, the performance metric can be :**   * **The RSRP difference between predicted L1-RSRP of Set B and measured L1-RSRP of Set B.** |
| ETRI | We support Option A.  For Option D, we cannot understand the direct relationship between the metrics and performance of AI/ML model. It may be helpful to indicate exact meaning of probability information. |
| ZTE | For Option A, the UE is required to report both the prediction results from model inference and measurement results from indicated Set A to the NW, which may increase UE reporting overhead. Option D only reflects the relative strength of different beams at one model inference, but lacks sufficient evaluations on how to justify the model performance only based on the probability information.  As far as we are concerned, the beam prediction accuracy related information in Option B has been thoroughly evaluated in Rel-18 and should be taken as the baseline for performance monitoring. Specifically, within a time duration or a monitoring window, the measured beam information based on indicated Set A and predicted beam information based on model output are compared and the beam prediction accuracy related information (e.g., Top-1/k beam prediction accuracy) is derived at the UE side. Compared with Option A and Option D, Option B can directly reflect the model performance and achieve reduced reporting overhead. Therefore, we suggest to take Option B as baseline and FFS other options.   * Support Option B: Report the beam prediction accuracy related information, take Top-1/K beam prediction accuracy as baseline |
| Panasonic | We think performance monitoring metrics should be discussed and concluded firstly, then we can discuss type 1 performance monitoring.  Regarding proposal 2-1A, we support Option A.   * For Option D, there could be the case that predicted beam is not accurate but probability information of predicted beam is still high because the model does not work well. Hence, we do not support Option B.   For other options, it is okay to put them as FFS. |
| Xiaomi | We support Option A, B and C.  For Option B and C, we have provided some examples in our tdoc. The definition of the performance metric need to be specified of course. For Option C, what does ‘assumption predicted RSRP already been reported to NW. and it is very hard to define the resource/method to obtain grand truth’ mean?  For Option D, it doesn’t need to be discussed since the probability information is not supported as inference results. Even it was supported as inference results, how does it work standalone for performance monitoring, does it mean it works on top of Option A? |
| Intel | First, we would like to second the idea that we should first focus on the metrics – that could help the discussion on the reporting easier.  Regarding the proposal itself, while we can support Option A – it may not be strictly necessary for UE to report the measurements for all beams in set A for this case; for instance, limited to set B of beams as mentioned by CATT. On the other hand, in that case, as FL also observed, Option C could be very similar to Option A, and in fact, preferable. Thus, if choosing between Options A and C, we should prioritize Option C.  Similar to MTK, QC, others, we also think Option B should be supported.  Lastly, similar to many others, we do not need Option D (probability) is a good metric for reporting for monitoring purposes. The probability information cannot convey model performance in general as there is no direct relation to the ground truth. Also, even if it could (in some cases), there isn’t a clear way to factor in any margin in the metrics (unlike L1-RSRP, etc.) to address the critical fact that the quality of two beams may be very similar to each other and just focusing on picking the Top 1 beam may not be an appropriate evaluation of the model performance. |
| New H3C | OK |
| InterDigital | We support Option A, but it is too early to take Option D. Option D needs further discussion. |
| Ericsson | **Option A**, there is no information of the predictions. Is the common understanding that it is already provided?  **Option D**. This is only relevant if the UE would have provided the probability information during the inference stage. This should be agreed first before agreeing on option D. Moreover, would not the UE also need to report the beam index of the strongest in this example?  **Option B+C**. Main text is ok with some minor wording changes, rest could be part of the FFS. |
| KDDI | Support Option A. Option D needs additional discussion as the relationship between probability information and monitoring performance is unclear. Option C is an option for Option A. Reporting RSRP differences may reduce the range of representation, reduce bit width, and reduce overhead. |
| SPRD | For Option A, both the prediction results from model inference and measurement results will be reported by UE. The overhead brings to UE cannot be ignored.  For Option D, there is no consensus whether the probability information is supported for inference. And it is uncertain whether this information can bring beneficial effects to monitoring. So we think we can wait until these two issues are resolved before discussing this option.  For Option B: Beam prediction accuracy can indeed characterize the quality of prediction performance, but it may require multiple calculations of prediction results to obtain this metric.  For Option C: UE can report the RSRP difference between predicted top-k beam and measured top-k beam. Compared to option A, this will greatly reduce reporting overhead.  For Option E: Similar with option D. |
| LG | We support option A for type 1 performance monitoring. However, we don’t think the sub-bullet (Strive for a common design as the report for NW-sided model) is needed, since we don’t think there is specification impact for performance monitoring of NW-sided AI/ML model.   * Support (Option A) to report the measurement results (e.g. L1-RSRP and/or beam information) of one set of beams, configured by NW   + Using existing CSI framework for configuration of the set of beams as the starting point     - FFS on necessary change for the report of measurement results: e,g, beam information only, L1-RSRP only, beam information and L1-RSRP     - ~~Strive for a common design as the report for NW-sided model~~   + Note: this may or may not have additional specification impact   We also support Option D. In this case, we need to further consider whether the probability information is per model/functionality, per report, per time instance, or per report parameter |
| Fujitsu | We support Option B and Option C.  Regarding Option A, in order to obtain the ground truth data, the UE needs to report a lot of measurement results which lead to huge amount of overhead.  For Option D, how this probability information works for performance monitoring requires more discussion. Since there is no comparison with ground truth data, whether this metric can actually reflect the performance is questionable. |
| Google | Support Option A. We think the overhead including the RS overhead and report overhead can be a potential issue to be studied.  We think other options can be FFS. |
| CMCC | We support option A and B.  For Option C, it is difficult to determine how much the RSRP difference can say the model performance is good.  For Option D, it is difficult to find the ground truth and define the metric. |
| CAICT | We are fine to focus on option A and D first and support Option A |
| Lenovo | We support option A and option C. |
| Fraunhofer | We support option A and D. |
| OPPO | We support  Option A (for Type 1 Option 1, NW monitoring)  Option B (for Type 1 Option 2 UE monitoring) and  Option D (probability of model output). |

#### Issue #2: Event to trigger the report

* Event-1: 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.*
* 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?*
* Event-3: The probability information of Top-1 or Top K beam of Set A is lower 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.*
* 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*

##### **(High)Questions to answer**

A: Whether to define events for report?

B: what else events in your mind?

|  |  |
| --- | --- |
| Company | Comments |
| HW/HiSi | On Question A: In our view it is not needed to define an event in case that a metric is reported to the NW. One of them should be sufficient in our view. We would like to hear the view of the FL and other companies whether both metric and event would be needed?  We firstly we should align our understanding what an event means, is it similar to handover events and then discussed by RAN2, or is its meaning something else?  Below our thoughts on the potential event candidates raised by the FL:   * Event-1: 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   + *=>* For a single failure, this does not seem to work. Some filtering would be needed. Also, if we would go this way, more discussion on the wording would be needed to make it unambiguous, what would “are different” man in detail, e.g. the same Top-K but different power ranking, some overlap between the predicted and measured Top-K, or completely orthogonal? * Event-2: The measured L1-RSRP of one set of beams is lower than a threshold.   + *=>* Not sure how this would be beneficial for monitoring. The measured beams can be different from the predicted. Also some measured beams maybe be larger than a threshold but are not close to the predicted beams in their direction. Or maybe all beams are below a threshold. * Event-3: The probability information of Top-1 or Top K beam of Set A is lower than a threshold   + => Not sure if this will work well. The probability of the predicted beams can be high according to the model’s inference. But the model may still not be working well, if it predicts false best 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   + => We need clarification, for our understanding: This may trigger a wrong event. The predicted RSRP and measured RSRP may differ. That is not so important for that the model is performing well. It is on the other hand important that the ranking among measured beams and the ranking among predicted beams is the same or similar. This event, does not give such an information. |
| TCL | Start a counter, if the number of RSRP less than a threshold for X times within a small-time window, then counter +1. If,   * the counter value reaches K within a larger-time window, * or within consecutive L small windows the counter +L, then the event occurs. |
| Vivo | A: Yes, events can be discussed. As deciding whether to define events relies on the methodology of the monitoring procedure, we prefer to discuss the performance metrics and the related report content firstly. In our view, if a performance matric is lower than a threshold, or fallback condition is satisfied can be candidate event. |
| MediaTek | Yes and support statistical results for at least Event 1, 3, 4, and support defining monitoring windows (number of samples, measuring frequencies,… etc) for these events. |
| NTT DOCOMO | A: Events can be defined for reporting.  B: We think the following event should be considered as well.  ・RSRP difference between measured [L1-]RSRP of current beam and predicted RSRP of the predicted Top 1 beam is larger than a threshold. |
| QC | OK with the direction of defining events, but the definitions will be highly dependent on the metrics that we define for performance monitoring. |
| CATT | The motivation of event-1 is not clear. If one of predicted Top-K beam is ideal Top-1 beam, does it mean the performance of AI/ML model is good enough? In our understanding, the performance of the AI/ML model is good although the predicted Top-K beams are not totally same with the ideal Top-K beams. We don’t see the motivation to support event-1.  For event-2, if the low L1-RSRP of beams of one time instance is due to obstruction of moving object, it can’t reflect the poor performance of AI/ML model. We suggest to update as；   * Event-2: The number of times the measured L1-RSRP of one set of beams is lower than a threshold.   For event-3, we are not sure the output of a model can be used as monitor a model performance. For example, if the performance of a AI/ML model is poor, how to ensure the output of AI/ML model (the probability information of Top-1) can be used for performance monitoring correctly? |
| ZTE | A: Yes. The event can be defined based on an indicated threshold from the NW.  B: It seems that all above events are only based on one model inference. Per our understanding, a single failure of model inference is not sufficient to determine whether the model is invalid or not. Considering the ping pong effect in wireless communication, the performance of the AI/ML model may vary over time, and inference errors can occur even for a valid AI/ML model. Therefore, to avoid unnecessary model switching, the model monitoring should be based on multiple observations of model inference performance over a period of time rather than a single observation. Therefore, we suggest to define the event based on statistic results of multiple observations. Besides, the procedures defined for beam failure recovery can be reused as much as possible, e.g.,   * events defined based on a counter on the number of failures (e.g., the predicted Top-1 beam is not the measured Top-1 beam) and a timer. |
| Xiaomi | First we need to clarify which type of performance monitoring is targeted for? Type 1 Option 1 or Type 2 Option 2?  We prefer to define event for Type 1 Option 2. While for Type 1 Option 1, we don’t think the event is needed since the event can be only be triggered by the calculated metric of UE. But if the UE has calculated, why not to report the calculated metric directly?  In addition, Event#5 can also be supported.   * Event-5: The predicted L1-RSRP difference between the measured L1-RSRP and the predicted L1-RSRP of the predicted Top-1 or Top K beam(s) of Set A |
| Intel | **A:** Yes, defining event-based reporting can help significantly in reducing ignaling OH.  **B:** On the particular events, we agree with QC that this depends on the decision on monitoring metrics.  On the list of events, we think Event-2 can be different from BFD – e.g., the thresholds can be different and the point is to not rely on things getting to a stage when BFR kicks in. Further, some kind of time-filtering (averaging)/consideration of multiple “sub-events” as suggested by CATT could be considered.  Further, we share view with CATT that Event-1 is not clear.  Lastly, we do not support Event-3 for similar reasons as for previous proposal. |
| NEC | Question A: yes  Question B: we think the following events should be considered.  1. based on the system performance, for example, if RLF or BFD is detected, then check if the failure is caused by the AI/ML.  2. one UE may be able to monitoring the beams which configured to other Ues for data transmission. |
| InterDigital | Question A: We support.  Question B: We believe further discussion is needed. Option A is not clear. We can support Option D. |
| Ericsson | We should first agree on what metrics to support, then we can further discuss whether an “event” is needed for such metrics. Also, if probability information is provided as part of the inference, we don’t see a need for event 3. Similarly, if RSRP is part of the prediction, this can mitigate the need for event 4.  **In summary,** first discuss performance metrics and inference result reporting (whether to support RSRP confidence and prob. Information). |
| KDDI | A: Yes. Events can be defined |
| SPRD | A: We would like to confirm if this is the previous step of issue 2 above. And is there a one-to-one correspondence between the event and report content in issue2? If so, we think we can merge these two issues for discussion. |
| LG | Question A: Yes, we think event-based performance monitoring can achieve reporting overhead reduction, since there is no periodic reporting of performance monitoring.  Question B: We support event-1 and event-4. Regarding these events, counter-based reporting within a certain time window can avoid ping-pong issue. |
| Fujitsu | We think the discussion on this proposal could be postponed. Agree with other companies that there is no need to report a single failure instance.  It’s better to have discussion on this after the option for Type-1 monitoring is agreed. |
| Google | We think event 2 and event 4 could be considered. |
| CMCC | A: Yes for Type 1 option 2 monitoring.  B: Event 1 with statistical results in a given window is reasonable, the details on the statistic method of prediction accuracy needs further discussion. The definition of Top K/1 can be a starting point. |
| CAICT | We are fine to define some events for report for AI/ML performance monitoring. Event 1 and 3 are preferred for further study. |
| Lenovo | Q-A: We think event based beam report for UE side performance monitoring is needed because the UE may have more knowledge on its model.  Q-B: At least for the case that predicted RSRP is available for the model output, the performance monitoring can be performed based on the differential RSRP for a certain set be beam pairs. |
| Fraunhofer | Q-A: Yes, we believe event-based reporting is required to ensure a good performance. |
| OPPO | A. Yes, define event(s).  B. We could start from Event-1, i.e. the beam prediction accuracy based event, and open to other event(s). |

#### 2.4.1.2 Type 2 performance monitoring for UE sided model

#### Issue #1: What to report from UE to NW on the operation

* Decision A: Report fallback operation from UE.
  + *Comments from FL: may be make sense. But do we need to specify the metric? Which metric?*
* Decision B: Report model selection/activation/ deactivation/switching
  + *Comments from FL: Do we need to report such operation? What NW can do with such report? It is hard for me to understand this report ☹*

#### Issue #2: Assuming associated ID is used for consistency, whether/how to validate an AI model?

* Method #1: Define a performance metric (RAN 1 or RAN 4) for model validation:
  + By offline test (e.g., up to RAN 4) If the performance can be achieved, UE can declare AI model is valid for a certain condition (including additional condition implied by associated ID.)
  + *Comments from FL: RAN 1 may be the right WG to define a performance metric for model validation. Whether/how to test can be left to RAN 4*
* Method #2: Configure a performance metric for model validation:
  + Real time validation (RAN 1 signaling): If the configured performance can be achieved, UE can activate the AI model. If the configured performance cannot be achieved, UE cannot activate the mode, or fallback from AI to non-AI scheme.
  + *Comments from FL: Fall back is needed, and there should be a controllable metric configured by NW. This can be used as “real time performance monitoring” (See Ericsson’s figure). That is, NW will configure a threshold of a metric to UE, if the metric cannot be met, fallback to non-AI. We can further discuss whether UE needs to check with this metric before activating a AI model (assuming after activation, UE is configured with resources for monitoring), if yes, how? If No, it means that UE will directly use a AI model as long as the associated ID is the same.*
* Method #3: Up to UE implementation. No specification in RAN 1.
  + *Comments from FL: assuming associated ID will be introduced.*

##### **(High)Questions to answer**

A: Whether need to report “selection/activation/ deactivation/switching” by UE, why?

B: Whether need to report “fallback” to non-AI by UE?

C: Whether need to define a procedure for “validation” and/or “activation” of AI model?

D: if B or C is yes, whether need to define metrics /events for validation/fallback?

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Company | A | B | C | Comments for A~C and D |
| FL | No | Yes | Yes | Event 1~ Even 4 or others can be FFS. |
| Hw/HiSi | No | [Yes] |  | B: Could be helpful for NW to know that UE has left AI-mode |
| TCL | Yes | Yes | Yes |  |
| vivo |  |  |  | The necessity of supporting type 2 performance monitoring is unclear |
| MediaTek | [Yes] | Yes | [No] | Question A: If the “selection/activation/ deactivation/switching” is reported as an “associated ID” change, then NW can adjust its NW additional condition accordingly. This way, UE does not need to report its preference on SetA/SetB configuration according to the new model.  Question B: The resource/report configuration of AI/non-AI can be different, so NW need to know if UE fallbacks to non-AI  Question C: Our answer is no for RAN1 procedure for validation. Since it is type 2 performance monitoring, “the validation” can be up to UE implementation. However, we think Method#1 is a valid method and should be defined in RAN4  Question D: Event needs to be defined for “fallback”, no event/report is needed for “validation” |
| NTT DOCOMO | No | No | Yes (w/ comment) | B: In our view, fallback operation is a switch from predicted beam reporting to measured beam reporting. Whether the predicted value is derived from AI or non-AI algorithm does not matter from NW perspective.  C: We think validation should be made for beam prediction reporting functionality not for AI model, unless model ID is introduced. In our view, Method#2 is more useful to guarantee the performance in the actual field. However, activation/deactivation should be made by NW, and only what UE reports should be only whether performance is met or not. |
| QC | No | Yes | Yes |  |
| CATT | [No] | - | - | Before answer FL’s question, we’d like to make sure whether Type-2 performance monitor is only applied to model-based LCM operation within same functionality? Since the situation is that only functionality-based LCM has been supported, we think model-based LCM is out of rel-19 scope.  If the Type-2 performance monitor is only applied to model-based LCM operation within same functionality, we think the answer of A is NO.  **And the UE can’t fallback to non-AI without gNB configuration** since the functionality should be configured by gNB. |
| ETRI | No | No | Yes | Question C: We agree with DCM’s comment. |
| ZTE | No | No | depends | A: Similar with the current procedures for beam failure recovery, the UE can monitor the performance of inactive functionalities (if possible) and report candidate functionality that can meet certain performance target. Such information is beneficial since the NW can perform subsequently functionality selection/activation/deactivation/switching efficiently based on the reported information. Anyway, the decision on selection/activation/ deactivation/switching is made by the NW.   1. No. The UE only need to report whether the currently monitored model is workable or whether new functionality is recommended, the decision on fallback is up to NW. 2. Fine to define a procedure for “validation” and/or “activation” of AI functionality (instead of model). Even when the NW-side additional conditions are aligned across training and inference, communication environments can vary dynamically and unpredictably over time. In such scenarios, performance monitoring-based approaches become crucial for functionality selection and validation at the UE side. 3. Yes. The procedures for performance monitoring can be reused as much as possible. |
| Panasonic | No | Yes | Yes | It may need to define metrics/events for validation/fallback. |
| Xiaomi | [Yes] | Yes | Yes | A: At least implicit report is needed, e.g., request the resource of a new set B  C: performance monitoring for non-activated AI model can be supported  D: Agree with FL |
| Intel | No | Yes | Yes |  |
| NEC | Yes | Yes | Open | Question D: Event for fallback is needed. And it requires NW to provide associated configuration for non-AI operations for the UE. |
| New H3C | NO | Yes |  |  |
| InterDigital | Yes | Yes | Yes | A: For better operation of BM, gNB and UE need to share the same understating on whether/how to use AI/ML model. In that sense, the reporting is beneficial.  B: AI/ML based BM is not always better and can be unreliable. In those cases, utilization of legacy BM can be beneficial.  D: Definition is needed. |
| Ericsson | NO | Yes, w. Mod.  [UE ~~report~~ request] | FFS | B. This should be handled via NW re-configuration.  The UE can hence “request” fallback to non-AI mode.  C. Mainly if real-time monitoring is needed. To be studied. |
| SPRD | No |  | No | Question C: It can be up to UE implementation. |
| LG |  |  |  | We have the same view as CATT, only functionality-based LCM has been supported. |
| Fujitsu | Yes | Yes | [Yes] | Since this is for Type-2 monitoring, we think it’s reasonable that UE could make decision and report the decision to the network, including selection/activation/ deactivation/switching/fallback.  Regarding C, one thing to clarify: is the “a procedure for validation and/or activation” different as performance monitoring? |
| Google | FFS | FFS | FFS | We think all of these can be FFS. |
| CMCC | [Yes] | [Yes] | [Yes] | If only Type 1 monitoring is supported for function based LCM, we do not need to discuss question A,B,C,D since decision is up to NW.  If Type 2 monitoring with UE report is supported for model based LCM, reporting “selection/activation/deactivation/switching/fallback” by UE is needed since new model or non-AI may need new configuration. |
| CAICT | No | Yes | [Yes] | Event 1/3 could be considered as starting point. |
| Lenovo | Yes | Yes | FFS | Q1: It should be needed for the NW to configure proper model input for inference and to align the understanding on the reported beam which may be based on the model output.  Q2: At least for the case that there is no available resources for model inference and the UE perform the non-AI operation, it should indicate this info to the NW. |
| Fraunhofer | No | Yes | Yes |  |
| OPPO | Yes | Yes | Yes | We think RAN2 had made the following agreement with highlighted yellow, saying UE has to report its decision to NW.  Agreements:  For UE-sided model, for the functionality management, the “network decision, network-initiated” AI/ML management is supported as a baseline. The following can be considered further “UE autonomous, decision reported to the network”, “Network decision, UE-initiated” (i.e. proactive approach).  “UE-autonomous, UE’s decision is not reported to the network” is not considered for Rel-19 |

#### 2.4.2 Performance monitoring for NW sided model

##### **(High)Questions to answer**

FL’s plan: low priority. Discuss the beam report, considering performance monitoring for NW sided model.

|  |  |
| --- | --- |
| Company | Comments |
| FL | Do we need to specify the metrics and new report for NW sided model?  If yes, what needs to be specified?  In FL’s view, we can discuss this together with the “measurement report” enhancement. What do you think? |
| HwHiSi | Metrics do not need to be specified for the NW sided model.  NW monitoring should be up to implementation. The same L1-RSRP reporting as for inference can be adopted. The UE does not need to know the purpose of the L1-RSRP reporting. |
| TCL | We don’t need. |
| vivo | It was demonstrated in our contribution that significant UCI overhead reduction can be achieved by UE-assisted model monitoring compared to direct measurement results report for NW-side AI/ML model. Further, similar framework of UE assisted performance monitoring for UE-side model can be reused to NW side model performance monitoring procedure.  Thus, we proposal,  For NW-sided model, for both BM-Case 1 and BM-Case2, support UE assisted performance monitoring procedure,   * FFS on detail metrics, including:   + Option B: Report the beam prediction accuracy related information   + Option C: Report the RSRP difference information between the measured and predicted   + Option D: Report probability information of the predicted beam of Top 1 or Top K beams   + Option E: Report confidence information of the predicted RSRP error   + Option F: Report hypothetical BLER-like metrics * FFS on whether to define event(s) to trigger above report(s)   Strive for a common design as the report for NW-sided model |
| MediaTek | Agree with FL, we can discuss this in the “measurement report” enhancement. |
| CATT | It seems not needed. The report for NW-sided model can be covered by ‘measurement report” enhancement’. For example, if the performance metric is Top-K beam prediction accuracy, the network can configure UE reporting the RS information of Top-K measured beams within Set A beam. |
| ETRI | We think that performance metrics do not need to be specified for NW-sided model.  However, measurement report enhancement should be discussed for overhead reduction during performance monitoring. Surely, it can be discussed together with other purpose such as model inference. |
| ZTE | No need to specify the metrics and new report for NW sided model |
| Panasonic | We are ok to discuss it later. We can discuss it together with the “measurement report” enhancement as FL’s comment. |
| Xiaomi | For performance monitoring, there is one point need to be specified: the report for inference should be reported by L1 but the report for performance can be reported by RRC. In this case, the mapping of the measurement results to each inference time instance should be indicated to NW for performance metric calculation. |
| Ericsson | Agree with FL |
| KDDI | Agree. Monitoring of NW-sided model can be implementation-dependent. |
| SPRD | We think there is no need to specify the metrics and new report for NW sided model. |
| LG | We don’t think there is spec impact. |
| Fujitsu | The beam report may need to be enhanced, considering the performance metric for monitoring.  Also, the reference signal configuration may need to be discussed for NW-side monitoring. |
| Google | We think it is necessary to consider the assistance info from UE side for NW-side monitoring, since only UE knows the measurement results for each beam. |
| CMCC | No need to specify the metrics and new report for NW sided model. |
| CAICT | Same understanding with FL. |
| Lenovo | No enhancement is needed. |
|  |  |

## 3 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** |

### 3.1 Report for inference

|  |  |
| --- | --- |
| Company | Comments |
| Futurewei | Proposal 2: For Rel-19 AI/ML-based BM, at least for inference for network-sided AI/ML model of BM-Case1 and BM-Case2, support reporting L1-RSRP(s) and its corresponding CRI/SSBRI(s) as beam related information in L1 signaling. |
| Ericsson | Proposal 19 For NW-sided model inference, regarding the FFS on max number of reported beam related information in one report, use 256 beams as a starting point (to support BMcase2 with setA equal to setB)  Proposal 20 For NW-sided model inference, regarding the FFS on report content for beam related information, use L1-RSRP, CRI (if needed) and SSBRI (if needed) as a starting point  Proposal 21 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.  Proposal 22 For NW-sided model inference, support methods for UEs to compress the set B temporal domain measurement results to reduce the reporting overhead. |
| Huawei/HiSi [3] | Proposal 4: For NW-sided model, the content for training, FFS  • Opt 1: Top M beam information of resource set(s) for Set A (No L1-RSRP)  • Opt 2: L1-RSRPs and beam index of Top M beam of resource set(s) for Set A  o FFS on the maximum value of M and how to determinate M, e.g, configured/predefined value/ according to a threshold/predefined method/etc…  • Opt 3: all L1-RSRPs of a resource set (without beam information or with best beam index (for differential L1-RSRP reporting))  • Note: Purpose, such as above “For NW-sided model, the content for training”, will not be specified in RAN 1.  Proposal 9: At least for NW-side model, for ~~further study~~ the reported beam information, at least support Opt 0 and further study other listed options   * Opt 0: legacy CRI/SSBRI, (i.e., index of resource in a resource set) * Opt 1: beam indexes are reported based on a bitmap, where bitmap indicating RS index of a resource set. * Opt 1-1: No additional beam index information required   + Note: This is used when L1-RSRPs are not reported for indicated bitmap. * Opt 1-2: The beam index with largest measured value of L1-RSRP is additionally reported.   + Note: This can be used when L1-RSRPs are also reported for indicated beams in the bitmap * ~~Opt 2: No beam index reporting.~~    + ~~Note: This can be used when L1-RSRPs are reported for all resources in a resource set.~~ * Opt 3: Only the beam index with largest measured value of L1-RSRP is reported (i.e., index of resource in a resource set)   + Note: This can be used when L1-RSRPs are reported for all resources in a resource set with differential L1-RSRP reporting; or when only Top 1 beam index is reported without L1-RSRP * ~~Opt 4: The beam index with largest measured value of L1-RSRP, and a bitmap are reported, where bitmap indicating RS index of a resource set,~~    + ~~Note: This can be used when L1-RSRPs are reported for indicated bitmap and/or beam index with largest measured value of L1-RSRP.~~ * ~~Opt 5: Index of a group of beams (identified as subset resource set of a resource set)~~   + ~~Note: This is used when all L1-RSRPs of the group of beams are reported.~~ * Opt 6: Adaptive selection among above options based on configurations of size of resource sets(s) and number of reported beams. * Other options are not precluded.   Proposal 17: For NW-sided model, for inference, the “beam related information” in a beam report, FFS   * Opt 1: L1-RSRPs and beam information of Top M beam of a resource set   + FFS on the maximum value of M and how to determinate M, e.g, configured/predefined value/ according to a threshold/predefined method/etc… * Opt 2: all L1-RSRPs of a resource set (without beam information or with best beam index (for differential L1-RSRP reporting)) * Opt 3: Index of a group of beams (identified as subset resource set of a resource set) ~~and all L1-RSRPs of the group of beams.~~   + ~~FFS on more than one group of beams~~ * FFS on other necessary information for BMCase-2 * FFS on the beam information * Note: Purpose, such as above “For NW-sided model, for inference”, will not be specified in RAN 1 specifications |
| Intel [4] | Proposal 9: 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 12: For a NW-sided AI/ML model, the beam report should contain the L1 measurement quantity (e.g. L1-RSRP) and the index of the measured RS.  Proposal 13: For a network-side AI/ML model, for BM-Case 2, 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 should provide L1 reports with measurements associated with related timestamps of measurements. |
| New H3C [5] | Proposal 3: For the NW side model, in the beam inference phase, use the same method for Set B measurement configuration as in the beam training phase.  Proposal 4: For the NW side model, in the beam inference phase, for BM Case 1, the maximum number of reported Set B beams in the L1 signaling should not exceed one quarter of the number of Set A beams. For BM Case 2, the maximum number of reported Set B beams in the L1 signaling should not exceed the number of Set A beams.  Proposal 5: For the NW side model, if the number of beams M measured in Set B exceeds the maximum number N allowed for reporting, only report the beams among these M measurements with the top N RSRP values. |
| Spreadtrum [7] | *Proposal 7: Reporting multiple past time instances in one reporting instance for BM-Case2 is not needed.* |
| Samsung [8] | 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 |
| Vivo [9] | Proposal 9: Report content supported in current specification can be re-used for data collection procedure with NW-side model. Additionally, time stamp information can be reported for BM-Case2.  Proposal 37: For model inference with NW-side model, support time stamp information as beam content for BM-Case2. |
| Interdigital [11] | ***Proposal 10:*** *Information about the time stamp for measurement instances should be supported.*  ***Observation 12:*** *For BM-Case 2 with a gNB-side AI/ML model, information about time stamp on measurement instances for reported beam related information needs to be reported for time domain prediction at gNB.*  ***Observation 14:*** *No need to restrict purpose of UE beam reporting, especially for network sided model, as how to utilize reported information is up to gNB implementation.* |
| CATT [11] | **Proposal 4: For NW sided model for BM case-1, for inference, the following options can be supported for the beam related information in a beam report:**   * **Opt 1: L1-RSRPs and beam information of Top M beam of resource set(s) where the value of M is configured；** * **Opt 2: all L1-RSRPs of resource set(s), and beam information of largest L1-RSRP from the RS of Set B beams.** |
| CMCC [14] | **Proposal 8: For NW-sided model, the following options are considered for a beam report at least for inference:**   * **Opt 1: All L1-RSRP from the resources for a set of beams with beam information**   + **where the set of beams are for UE measurement and report**   + **FFS on how to report beam information** * **Opt 2: A subset of L1-RSRP from the resources for a set of beams, with beam information**   + **where the set of beams are for UE measurement and report**   + **FFS on how to determinate the subset, at least including data selection, data omission** |
| LGE [18] | **Proposal #5. Regarding the report of more than 4 beam related information in L1 signaling, support CRI/SSBRI+L1-RSRP for the report content as legacy.** |
| Panasonic [19] | **Proposal 1: NW-sided model inference, support to that a measurement window can be configured with the measurement resource set.**  **Proposal 2: NW-sided model inference, support to option 1. In addition, option 2 can be supported if the exact subset selection of best *K* beams is up to UE.**  **Proposal 3: Group-based beam reporting is modified to support to report more than 4 beams in one report.** |
| GOOGLE [23] | ***Proposal 3: For beam report for NW model inference for SD beam prediction, UE reports the following information:***   * ***L1-RSRP for a set of SSBs/CSI-RSs 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*** |
| ZTE [24] | ***Proposal 12: 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 13: 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.*** |
| CAICT [25] | ***Proposal 4: FL’s proposal on NW-sided model reporting content for inference is acceptable.*** |
| **Fraunhofer [30]** | **Proposal 1: For NW-sided models, for inference, the UE should report at least the L1-RSRP in a beam report.** |
| **Sharp [33]** | **Proposal 6:** For NW-sided model, for BM-Case 2, support of measurement results of multiple past time instances in one CSI report is not necessary. |
| **MediaTek [34]** | ***Proposal 1: 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 2: For NW-side model inference, same design of report and resource configurations can be used for BM Case1 and BM Case2***  ***Proposal 3: For NW-side model inference, study both periodic and aperiodic report and Set B resource configurations.***  ***Proposal 5: For NW sided model, for inference, consider following options for “beam related information” in a beam report,***   * ***Opt 1: L1-RSRPs and CRI/SSBRIs of Top M beam of the configured resource set***   + ***FFS on how to determine M*** * ***Opt 2: all L1-RSRPs of the configured resource set, without CRI/SSBRI or with only the best CRI/SSBRI (for differential L1-RSRP reporting)***   ***Proposal 6: For NW-side model inference, consider re-using the Rel-18 report sub-configuration framework for UE to report the measurement results of more than 4 beams in one report***  ***Proposal 7: For NW-side model inference, for >4 beam L1 reporting, further consider specification on the limitation of the reported beams in one report*** |

##### **Summary on report content for inference**

For NW sided model, for inference, the “beam related information” in a beam report, at last for BM-case 1, further study with potential down selection with the following option:

* Opt 1: L1-RSRPs and beam information of Top M beam of a resource set
  + FFS on the maximum value of M and how to determinate M,
    - Alt 1: reporting of at most N strongest set B beams
    - Alt 2: reporting of only beams within X dB of the strongest beam
  + *Supported by: Ericsson, Huawei/HiSi, CATT, CMCC, LGE, Panasonic*
* Opt 2: all L1-RSRPs of a resource set (without beam information or with best beam index (for differential L1-RSRP reporting))
  + *Supported by: Huawei/HiSi, CATT, CMCC, Panasonic*
* ~~Opt 3: Index of a group of beams (identified as subset resource set of a resource set) and all L1-RSRPs of the group of beams.~~
  + ~~FFS on more than one group of beams~~
  + *~~FL’s comment: this is part of beam information~~*
* FFS on other necessary information for BMCase-2
  + *Ericsson: support methods for UEs to compress the set B temporal domain measurement results to reduce the reporting overhead.*
  + *CATT: 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 instance in a pre-defined order in a beam report.*
* Note: Purpose, such as above “For NW-sided model, for inference”, will not be specified in RAN 1 specifications

##### **Summary on whether to support report measurements from multiple time instances in one report for BM-Case2, together with time stamp information?**

* **Spreadtrum (No/Yes):**Reporting multiple past time instances in one reporting instance for BM-Case2 is not needed (inference). For the NW side model, in the beam training phase, For BM Case 2, it is suggested to include the timestamp of the instance's measurement in the reported results.
* **Samsung (Yes):** 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
* **Vivo (Yes):** Report content supported in current specification can be re-used for data collection procedure with NW-side model. Additionally, time stamp information can be reported for BM-Case2.
* **Interdigital (Yes):** Information about the time stamp for measurement instances should be supported.
* **GOOGLE (Yes?)*:*** 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
* **Sharp (Yes):** For NW-sided model, for BM-Case 2, support of measurement results of multiple past time instances in one CSI report is not necessary.
* **MediaTek (Yes):** 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.
* **CATT (yes):** 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 instance in a pre-defined order in a beam report.
* **Nokia (Yes):** For 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 for BM-Case2 time stamp information can be derived from report based on report configuration.
* **KDDI(Yes):** For the L1-RSRP reporting of the NW-sided model in BM-Case 2, consider supporting the reporting of the difference over time.
* **Huawei/HiSi (No?):** 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 it would inflict.
* **ZTE (yes):** 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. 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.
* **LGE (Yes for Higher layer):** 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)
* **Xiaomi(yes):** 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.
* **Ericsson (Yes):**support methods for UEs to compress the set B temporal domain measurement results to reduce the reporting overhead.

##### **Summary on FFS on max number of reported beam related information in one report**

* Ericsson: For NW-sided model inference, regarding the FFS on max number of reported beam related information in one report, use 256 beams as a starting point (to support BMcase2 with setA equal to setB)
* ZTE: 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.
* Interdigital:For network sided model, support enhanced UE reporting to report up to 64 RSRP values for whole Set A over multiple time instances.

### 3.2 Report for monitoring and/or training

|  |  |
| --- | --- |
| Company | Proposal/observations |
| Ericsson [2] | Proposal 23 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, -> FL: Not so sure on what is this, not in the Tdoc  • 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 24 For NW-sided data collection, conclude that it is up to RAN2 on whether RRC/MDT procedures should be supported |
| Huawei/HiSi[3] | Proposal 3: From RAN 1 perspective, for NW-sided model, conclude that for monitoring and/or training, the report with more than 4 beam related information in L1 signaling can 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 for inference can be used for training and/or monitoring.  • Note: Purpose, such as above “For NW-sided model, for monitoring and/or training”, will not be specified in RAN 1. |
| Intel [4] | Proposal 8: 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. |
| H3C[5] | For the NW side model, in the beam training phase, For BM Case 2, it is suggested to include the timestamp of the instance's measurement in the reported results. |
| Spreadtrum[7] | ***Proposal 2：For data collection for training at NW-side, option 1 and option 2 can be considered.***   * ***Opt 1: UE only report L1-RSRPs from the resource in one measurement set*** * ***Opt 2: UE report L1-RSRPs from the resource in one measurement set and partial L1-RSRPs with resource index from the resource in other measurement set*** |
| Samsung[8] | **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.** |
| Vivo [9] | **Proposal 10: For data collection procedure with NW-side model, support to report UE measurement results via L1-layer signaling and higher-layer signaling.**  **Proposal 11: For data collection procedure with 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 12: For data collection procedure with NW-side model, it is crucial to investigate approaches to minimize the overhead of the report transmitted through L1-layer or higher-layer signaling.** |
| Interdigital [11] | ***Observation 14:*** *No need to restrict purpose of UE beam reporting, especially for network sided model, as how to utilize reported information is up to gNB implementation.*  ***Observation 15:*** *For network sided model, enhancement of UE reporting can be considered to allow both training and inference via UE beam reporting.*  ***Proposal 13:*** *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.* |
| **CATT [12]** | **Proposal 13: For NW-sided model, at least for BM-Case1, training data collection content should consider the following options:**   * **Opt. 1: L1-RSRP from resource of Set B and L1-RSRP from resource of Set A;**   + **For the case** **Set B is a subset of Set A, Opt. 1 can be L1-RSRP from resource of Set A** * **Opt. 2: L1-RSRP from resource of Set B and Top-K beam information among beams of Set A.** * **Signaling for training data collection**   **Proposal 14: For NW-sided model, at least L1 signaling can be considered for reporting the contents of training data.**   * **Whether to report the contents in one or multiple measurement report**   **Proposal 15: For NW-sided model, for the case Set A and Set B are different, the measurements from Set A and measurements from Set B can be conveyed via separate reports, respectively.**  **Proposal 16: For NW-sided model, for the case Set B is a subset of Set A, the measurements can be conveyed in one report.**   * **RX beam assumption** |
| **China Telecom [13]** | ***Proposal 1: For NW-sided model, at least for BM-Case1, all three options can be supported for training data collection content:***   * ***Opt.1: M1 L1-RSRPs (corresponding to M1 beams) with the indication of beams (beam pairs) based on the measurement corresponding to a beam set, where M1 can be larger than 4, if applicable*** * ***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 (beam pair) indices based on the measurement corresponding to a beam set, where M3 can be larger than 4, if applicable***   ***Proposal 2: For NW-sided model, at least L1 signalling can be considered for reporting the contents of training data.***   * ***FFS：Whether to report the contents in one or multiple measurement report*** |
| **CMCC [14]** | **Proposal 1: For NW-sided model, for data collection for training, at least for BM-Case1, option 1a and 2a are preferred for the contents:**  **Note: Purpose, such as above “For NW-sided model, for data collection for training, at least for BM-Case1”, will not be specified in RAN 1 specifications**   * **Opt 1a: L1-RSRPs based on one measurement resource set and beam information**   + **FFS on report all or a subset of L1-RSRPs from the resource set, at least including data selection, data omission, e.g., L1-RSRP(s) higher than a threshold**   + **FFS on how the corresponding beam information is reported.**   + **Where the set of beams includes Set A, or Set A +Set B** * **Opt 2a: L1-RSRPs based on one measurement resource set, and beam information for Top 1 (FFS: Top K) beam(s) based on another measurement resource set**   + **FFS on report all or a subset of L1-RSRPs from the resource set, at least including data selection, data omission, e.g., L1-RSRP(s) higher than a threshold**   + **FFS on whether/how the corresponding beam information needs to be reported or not.** |
| **LGE [18]** | **Proposal #1: For data collection, RAN1 to focus on inference aspects, and it is up to RAN2 to specify/enhance higher-layer based approach for data collection for training.**  **Proposal #2: 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)** |
| **Fujitsu [20]** | ***Proposal 1:***   * *Regarding training data collection for both UE side model and NW-side model, L3 signaling is preferred as the reporting container.*   ***Proposal 2:***   * *For training data collection, at least the following information should be included:*   + *Reference signal ID*   + *Beam quality, e.g., L1-RSRP*   ***Proposal 3:***   * *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 4:***   * *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 a subset of Set A if Set B is different from Set A.*   ***Proposal 5:***   * *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.* |
| **Xiaomi [21]** | ***Proposal 3-1: Report collected data for NW-side model training by RRC signaling.***  ***Proposal 3-2: 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 3-3: 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.*** |
| **NEC [22]** | Proposal 9: For report content for NW-side model training, Option 2 and Option 3 should be supported.  − Opt 2: L1-RSRPs and beam index of Top M beam of resource set(s) for Set A  − Opt 3: All L1-RSRPs of a resource set (without beam information or with best beam index (for differential L1-RSRP reporting))  Proposal 10: Support sub configuration of CSI report for AI/ML, each sub configuration may correspond to a different model or a different AI/ML LCM stage requiring data collection. |
| **GOOGLE [23]** | ***Proposal 2: For beam report for NW model training for SD beam prediction, UE reports the following information:***   * ***SSBRI/CRI with the best L1-RSRP measured*** * ***L1-RSRP for a subset of SSBs/CSI-RSs configured by the NW*** * ***Hypothetical measurement error for the subset of SSBs/CSI-RSs with L1-RSRP reported***   ***Proposal 4: For beam report for NW model monitoring for SD 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*** * ***L1-RSRP for a subset of SSBs/CSI-RSs configured by the NW***   ***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 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*** |
| **ZTE [24]** | Proposal 3: 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 |
| **CAICT [25]** | ***Proposal 5: For NW-sided model training, Opt.3 should be considered for high layer signaling rather than L1 signaling.*** |
| **ETRI [27]** | **Proposal 3: For NW-sided model, support an additional beam set for performance monitoring.**  **Proposal 4: Support** **methods for reducing UE reporting overhead during the data collection for training when the AI/ML model is located on NW-side.** |
| **OPPO [29]** | **Proposal 1: 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 2: For BM-Case2, the temporal domain information of collected data should be reported to NW in an implicit manner in L1 signaling (no explicit time stamps needed).** |
| Fraunhofer [30] | **Proposal 17: 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.** |
| Nokia [31] | **Proposal 16: Consider the following for a CSI report that enables beam prediction at the NW,**   * **For BM-Case1 and BM-Case2, when reporting beam related information for N>4 beams for each reporting instance, support**   + **Option 1: Beam related information can be N > 4 CRIs/SSBRIs and corresponding L1-RSRPs of one measurement resource set.**   + **Option 2: Beam related information can be an indication of resource subset and corresponding L1-RSRP of the resources within the resource subset. The UE may be configured with multiple resource subsets (e.g., can corresponding to fixed beam patterns considered by the NW) for reporting.** * **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: Value(s) of X (larger than legacy X>2dB) to reduce UCI reporting overhead.** * **For 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 specific beams/RSs can be reported for each reporting instance, including beam selection and/or beam omission.** * **FFS: whether for BM-Case2 time stamp information can be derived from report based on report configuration.**   **Proposal 19: 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.** |
| Sharp [33] | **Proposal 1:** For NW-side AI/ML model, support the following options as contents of training:   * Opt.1: Top M beam index of a measurement resource set * Opt.2: L1-RSRPs and beam index of Top M beam of a measurement resource set * Opt.3: all L1-RSRPs of a measurement resource set * Indication of which option is used for reporting is provided to UE.   **Proposal 2:** For NW-sided model, for training, in a beam report initiated by network, support the measurement result report of more than 4 beam related information in L1 signalling. |
| MediaTek [34] | ***Proposal 14: For the content of the reported training data, besides the*** ***measurement of Set B for model input data, consider the following options for model output data:***   * ***Opt 1: CRI/SSBRI of Top-M resources in Set A (No L1-RSRP)*** * ***Opt 2: L1-RSRPs and CRI/SSBRI of Top-M resources in Set A***   + ***FFS on how to determinate M, e.g, configured/predefined value/according to a threshold*** * ***Opt 3: all L1-RSRPs of a full/subset of Set A (without CRI/SSBRI or with best CRI/SSBRI for differential L1-RSRP reporting)***   ***Proposal 15: For the reporting of the collected training data, at least when the data content includes L1-RSRP, RAN1 prioritize the discussion on using higher layer signaling for reporting.*** |
| ITL [38] | ***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. FFS on further down-selection and data selection.***  ***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 9: For NW-side model inference, the purpose for which the gNB uses the measurements reported from the UE is transparent to the UE*** |
| CEWiT [39] | ***Proposal 3 For NW-sided model, for monitoring and training, the report for inference with more than 4 beam related information can be used.*** |
| KDDI [41] | ***Proposal 5：For the L1-RSRP reporting of the NW-sided model in BM-Case 2, consider supporting the reporting of the difference over time.*** |

##### **Summary on the views for report for monitoring or training**

(As conclusion)

From RAN 1 perspective, for NW sided model, for monitoring and/or training, the report for inference with more than 4 beam related information in L1 signaling can be used.

* *Supported by: Huawei/Hisi, vivo, CATT, CT, Interdigital (?)*

##### **Summary on the views for Content for monitoring or training**

For NW sided model, the content for training, FFS

* Opt 1: Top M beam information of resource set(s) for Set A (No L1-RSRP)
* Opt 2: L1-RSRPs and beam index of Top M beam of resource set(s) for Set A
  + FFS on the maximum value of M and how to determinate M, e.g, configured/predefined value/ according to a threshold/predefined method/etc…
* Opt 3: all L1-RSRPs of a resource set (without beam information or with best beam index (for differential L1-RSRP reporting))
* Note: Purpose, such as above “For NW-sided model, the content for training”, will not be specified in RAN 1.

##### **Summary on the views for container**

* **Intel:** inference/model monitoring data collection should use L1 based reporting, while data collection for training should use RRC or MAC-CE based reporting.
* **Samsung:** 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.
* **Vivo:** For data collection procedure with NW-side model, support to report UE measurement results via L1-layer signaling and higher-layer signaling.
* **CATT:** For NW-sided model, at least L1 signaling can be considered for reporting the contents of training data.
  + Whether to report the contents in one or multiple measurement report
* **China Telecom：**For NW-sided model, at least L1 signalling can be considered for reporting the contents of training data.
* **LGE**：Proposal #1: For data collection, RAN1 to focus on inference aspects, and it is up to RAN2 to specify/enhance higher-layer based approach for data collection for training.
* **Fujitsu:** Regarding training data collection for both UE side model and NW-side model, L3 signaling is preferred as the reporting container.
* **CAICT**: For NW-sided model training, Opt.3 should be considered for high layer signaling rather than L1 signaling.
* **Fraunhofer:** 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.
* **Sharp:** For NW-sided model, for training, in a beam report initiated by network, support the measurement result report of more than 4 beam related information in L1 signalling.
* **MediaTek:** For the reporting of the collected training data, at least when the data content includes L1-RSRP, RAN1 prioritize the discussion on using higher layer signaling for reporting.
* **ITL:** For NW-side model data collection for training, it is proposed to at least support the higher layer signaling to convey data collection contents

##### **Whether to configure Set A, Set B for NW sided model training and/or inference separately?**

* **CATT:** 
  + For NW-sided model, for the case Set A and Set B are different, the measurements from Set A and measurements from Set B can be conveyed via separate reports, respectively.
  + For NW-sided model, for the case Set B is a subset of Set A, the measurements can be conveyed in one report.
* **Fujitsu:** 
  + 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.
  + 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 a subset of Set A if Set B is different from Set A.

### 3.3: Overhead reduction and beam information

|  |  |
| --- | --- |
| Company | Proposals |
| Huawei/HiSi [4] | Proposal 6: For reporting overhead reduction of NW-side AI/ML model, regarding omission/selection of collected data, discuss the following options:  • Opt1: Only report Top M beams with highest RSRP.  • Opt2: Only report the beams for which the RSRPs are within a threshold to the strongest RSRP.  Proposal 7: At least for NW-sided model, the quantization of a reported L1-RSRP value,  • Support differential L1-RSRP reporting with legacy quantization step and range  o FFS: whether introduce new step size(s) and/or range(s) applicable to absolute of L1-RSRP and/or differential L1-RSRP  • ~~FFS on whether to support absolute L1-RSRP reporting (for all beams in a set)~~  ~~• FFS on whether to support reporting the normalized L1-RSRP measurement instead of actual L1-RSRP values~~  Proposal 8: 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 it would inflict. |
| TCL [6] | ***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.*** |
| Spreadtrum[7] | ***Proposal 4: For BM-Case1 and BM-Case2 with a network-side AI/ML model, existing quantitative criteria should be reused at least for model inference.*** |
| Samsung [8] | **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** * **Support CRI/SSBRI omission in case of the number of reported beam(s) is the same as the number of resources within the set for channel measurement** * **FFS: Whether to support new quantization method and CRI/SSBRI.** |
| Vivo[9] | **Proposal 13: For data collection procedure with NW-side model, support enhancements on quantization range and quantization step to reduce overhead for measurement results report.**  **Proposal 14: For data collection procedure with NW-side model, support adaptive number of beams in a beam report.**  **Proposal 15: For data collection procedure with NW-side model, considering 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 16: For data collection procedure with NW-side model, consider to use time domain data compression to reduce overhead.**  **Proposal 35: For model inference with NW-side model, support beam pattern indicator as report content to indicate a subset beams of a group of beams included in beam report.**  **Proposal 38: For model inference with NW-side model, support enhancements on quantization range and quantization step to reduce overhead for measurement results report.**  **Proposal 39: For model inference with NW-side model, support adaptive number of beams in a beam report.** |
| Apple [10] | **Observation 4-1: weak beams’ RSRPs can be omitted in the beam reporting for overhead reduction.**  **Observation 4-2:**   * **Reporting selected beams out of all set B beams rather than reporting all set B beams is beneficial in reducing feedback overhead, which can be supported by bitmap(s) or combinatorial index/indices.** * **Using a common reference beam across multiple occasions helps reduce feedback overhead.** * **Consider temporal correlation, the signaling of selected un-omitted beams can be shared among occasions, which can be supported by a common bitmap or a common combinatorial index across occasions.**   **Proposal 4-1: to control feedback overhead, beam reporting for BM Case-1 consists of**   * **Indication of the strongest beam index** * **Indication of the number of un-omitted beams** * **The strongest beam’s RSRP** * **Bitmap to indicate un-omitted beams** * **Differential RSRPs for uno-omitted beams except the strongest beam**   **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** * **Indication of the number of un-omitted beams** * **The strongest beam’s RSRP** * **Differential RSRPs for un-omitted beams except the strongest beam** |
| InterDigital [11] | ***Observation 9:*** *The UE’s frequent reporting of AIML input measurements for inference at gNB greatly increases UE reporting overhead.*  ***Observation 10:*** *Beam prediction accuracy with sparse reporting (e.g., reporting odd/even beams alternatively or best 50% of beams) is as good as baseline case (i.e., reporting all beams in Set B).*  ***Proposal 8:*** *Support a sparse reporting mechanism for data collection for inference (e.g., based on channel conditions).*  ***Proposal 9:*** *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.* |
| CATT [12] | **Proposal 6: At least for NW-sided model, for inference, support differential L1-RSRP reporting with legacy quantization step and range.**  **Proposal 7: At least for NW-sided model, for inference, support to introduce a larger quantization step size for differential L1-RSRP reporting.**  **Proposal 8: 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.** |
| Lenovo [16] | **Proposal 2: Support procedures that enable a UE to transmit a subset of the set of measured/collected samples from the environment (e.g., more informative samples among all samples).**  **Proposal 16: Study schemes on differential RSRP report for UCI overhead reduction for larger number of beam reports in a beam report.** |
| Fujitsu [20] | * *For BM Case-1 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 22:***   * *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 5:***   * *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.* |
| Xiaomi [21] | ***Proposal 3-3: 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 3-4: 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.*** |
| NEC [22] | Proposal 12: For overhead reduction, support omitting at least part of CRI/SSBRI information.  Proposal 13: 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. |
| ZTE [24] | Proposal 4: Regarding measurement results report,   If measurement results of all beams in a measured beam set are to be reported, support enhanced reporting methods for the purpose of reporting overhead reduction, e.g., beam ID can be obtained implicitly from the reporting order of all measured RSRPs.   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 overhead reduction, support the following two options of Method #1: omission/selection of measurement quantity for the beam report in L1 signaling for content(s) for one time instance in one reporting,  • Opt 1: Only report Top M beams with highest RSRP  • Opt 2: Only report the RSRP larger than a threshold or within a threshold  Proposal 8: For overhead reduction, support to specify threshold based reporting with configurable minimum and maximum number of reported beam related information in a single report.  Proposal 9: For overhead reduction, support specification enhancements for data omission among samples (e.g., according to data quality).  Proposal 10: For overhead reduction, support the following two options of Method #2: Reduce the overhead of beam information for the beam report in L1 signaling for content(s) for one time instance in one reporting,  • Opt 3: Only the beam index with largest measured value of L1-RSRP is reported. L1-RSRPs are reported for all resources in a resource set.  • Opt 4: The beam index with largest measured value of L1-RSRP, and a bitmap are reported, where bitmap indicating RS index of a resource set, and L1-RSRPs are reported for indicated bitmap and/or beam index with largest measured value of L1-RSRP.  Proposal 11: At least for NW sided model, for the quantization of a reported L1-RSRP value, support differential L1-RSRP reporting with new quantization step size applicable to absolute of L1-RSRP and/or differential L1-RSRP.  Proposal 14: 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 super set of beam IDs to be reported  Proposal 15: 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. |
| CAICT [25] | ***Proposal 6: Support FL’s proposal for quantization of a reported L1-RSRP value for NW-sided model and finer step size should be considered.***  ***Proposal 7: Support FL’s proposal for beam information in the report for NW-sided model and Opt.2 could be listed*** ***as a note rather than an option.*** |
| ETRI [27] | **Proposal 1: For NW-sided model, reducing measurement overhead is necessary for model inference, especially in the case of temporal domain prediction.**  **Proposal 2: For NW-sided model, in model inference, support the method of omitting RSRP values based on differences in RSRP values.** |
| OPPO [29] | **Proposal 3: For BM-Case1 with NW-side model, reduce the reporting overhead for both fixed and variable Set B, e.g. by dropping the part of SSBRIs/CRIs.**  **Proposal 4: For BM-Case2 with NW-side model, UE reports multiple measurement instances of Set B in a single beam reporting instance.** |
| Fraunhofer [30] | **Proposal 2: For NW-sided models, for inference, depending on the configuration the UE can omit at least a part of the beam IDs to reduce reporting overhead.**  **Proposal 18: Adopt report overhead reduction techniques other than the legacy Option 1.**  **Proposal 19: Consider the following approaches to enhance the quantization of differential RSRPs:**   * **Multi-resolution quantization,** * **Increased step sizes,** * **Adaptive reference beam for differential RSRPs.** |
| Nokia [31] | * **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: Value(s) of X (larger than legacy X>2dB) to reduce UCI reporting overhead.** |
| DoCoMo [32] | **Proposal 16: 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.**   **Proposal 17: Support subset beam reporting for variable Set B with pre-defined beam patterns for NW side model.** |
| MediaTek [34] | ***Proposal 4: At least for NW sided model, consider the following quantization options of a reported L1-RSRP value,***   * ***Option1: Differential L1-RSRP reporting***    + ***Whether to use legacy/new quantization step and range*** * ***Option2: Absolute L1-RSRP reporting (for all beams in a set)***   + ***Whether to use legacy/new quantization step and range*** * ***Option3: Normalized L1-RSRP measurement reporting***   + ***Whether/how to align the reference values for normalization between NW and UE*** |
| CEWiT [39] | ***Proposal 4 For NW-sided model, for inference, support Opt-1 and Opt-2, i.e., L1-RSRPs and beam information of Top M beam of a resource set and all L1-RSRPs of a resource set , for the beam related information.***  ***Proposal 5 For NW-sided model, for inference, support implicit indication of CRI or SSB-RI to reduce the size of report content for beam related information.***  ***Proposal 6 For quantization of reported L1-RSRP values, support introducing new step size and range for reporting differential L1-RSRP values.*** |
| KDDI [41] | ***Proposal 1:***  ***An option with a quantization step size larger than 2 dB and with a wider range should be supported.***  ***Proposal 2: Support only reporting the RSRP larger than a threshold or within a threshold.***  ***Proposal 3: Two methods should be considered for setting thresholds for adaptive reporting content reduction: pre-setting by rules or configuration from the network.***  ***Proposal 4: Methods to allow adaptive changes by the UE in the number of reporting beams for overhead reduction should be considered.*** |

Fraunhofer [30]

Table 1: List of report overhead reduction methods.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Scheme: | | | | Report size excluding CRC in bits  – number of measured/predicted beams  – number of reported beams  – number of patterns  – Random Variable (RV) representing the number of beams above the threshold |
|  | Selection | Quantization | Indexing |
| Opt1 | Top-M | Diff-RSRP | Beam-ID |  |
| Opt2 | Top-M | Diff-RSRP | Bitmap |  |
| Opt3 | Top-M | Diff-RSRP | Combinatorial |  |
| Opt4 | Threshold | Diff-RSRP | Beam-ID | ,  where |
| Opt5 | Pattern-P | Diff-RSRP | Pattern-ID |  |
| Opt6 | All | Abs-RSRP | No ID |  |
| Opt7 | All | Diff-RSRP | Single Beam-ID |  |

In Figure 2, we show the resulting report sizes of different report overhead reduction techniques over the increasing number of measured/predicted beams N.

|  |  |
| --- | --- |
|  |  |
| 1. M = 8 | 1. M = 16 |

Figure 2: Reporting overhead comparison utilizing different reduction techniques.

##### **Summary on views for quantization**

At least for NW sided model, the quantization of a reported L1-RSRP value,

* Support differential L1-RSRP reporting with legacy quantization step and range
  + FFS: whether introduce new step size(s) and/or range(s) applicable to absolute of L1-RSRP and/or differential L1-RSRP
  + *Supported by CATT, vivo, CAICT*
  + *Larger step:* 
    - *Yes: vivo, Samsung, Apple?, CATT, ZTE, Fraunhofer, Nokia, DoCoMo,* *CEWiT, KDDI*
    - *FFS: Huawei,*
    - *No: Spreadtrum?*
  + *Smaller step: High-resolution, non-* *differential: Fujitsu????*
* FFS on whether to support absolute L1-RSRP reporting (for all beams in a set)
* FFS on whether to support reporting the normalized L1-RSRP measurement instead of actual L1-RSRP values

##### **Summary on views for beam information in the report:**

At least for NW-side model, further study the reported beam information

* Opt 0: legacy CRI/SSBRI, (i.e., index of resource in a resource set)
* *Supported by Huawei CATT (and resource set id if multiple resource sets consists set B;), Fujitsu*
* Opt 1: beam indexes are reported based on a bitmap, where bitmap indicating RS index of a resource set.
  + Note: This is used when L1-RSRPs are reported for indicated bitmap.
* *Supported by Huawei, CATT*
* Opt 2: No beam index reporting.
  + Note: This can be used when L1-RSRPs are reported for all resources in a resource set.
* *Supported by CAICT as note*
* Opt 3: Only the beam index with largest measured value of L1-RSRP is reported (i.e., index of resource in a resource set)
  + Note: This can be used when L1-RSRPs are reported for all resources in a resource set with differential L1-RSRP reporting; or when only Top 1 beam index is reported without L1-RSRP
* *Supported by Huawei:*
* Opt 4: The beam index with largest measured value of L1-RSRP, and a bitmap are reported, where bitmap indicating RS index of a resource set,
  + Note: This can be used when L1-RSRPs are reported for indicated bitmap and/or beam index with largest measured value of L1-RSRP.
* Opt 5: Index of a group of beams (identified as subset resource set of a resource set)
  + Note: This is used when all L1-RSRPs of the group of beams are reported.
  + *Supported by vivo, Interdigital, CATT?*
* Opt 6: Adaptive selection among above options based on configurations of size of resource sets(s) and number of reported beams.
* *Supported by Huawei:*

### 3.4 1st Round discussion

#### Issue #1: L1 Report content for NW-sided model

For NW-sided model, ~~for inference,~~ the “beam related information” in a beam report in L1 signaling, at last for BM-Case 1 and for one time instance of BM-Case 2, support the following options:

*FL’s comments:*

* *Delete “for inference” is to extent the usage for monitoring and training*
* *“in L1 signaling”, for training, the content is to be discussed separately.*
* *“for one time instance of BM-Case 2”, to avoid repeatedly discuss this for BM-Case 2. It is only about the content. Format or some other optimization is a separate discussion.*
* Opt 1: L1-RSRPs and corresponding beam information of Top M beam(s) of a resource set
  + FFS on the maximum value of M and how to determinate M, FFS:
    - Alt 1: M strongest with highest L1-RSRP, where M is configured by gNB
    - *FL’s comments: this is legacy*
    - Alt 2: M beams within X dB gap to the highest L1-RSRP
  + FFS on the beam information
  + *FL’s comments: this can be used for inference, monitoring, data collection (may not be sufficient) via L1 signaling, Alt 1 and Alt 2 is part of the beam omission.*
* Opt 2: All L1-RSRPs of a resource set
  + FFS: without beam information or with best beam index (for differential L1-RSRP reporting, if supported))
  + *FL’s comments: This method can be used for inference, monitoring, data collection (of Set A and/or Set B). there is no beam omission.*
* FFS
  + *FL’s comments: FFS is further optimization for monitoring/training..*
  + Opt 3: Beam information of Top M beam(s) of a resource set
    - *FL’s comments: assuming measurement of Set B is in a separate report for training,*
    - *This can be used for monitoring*
  + Opt 4: Opt 3 for one resource set, and Opt 1 or Opt 2 for another resource set.
    - *FL’s comments: assuming measurement of Set B with Opt 1/Opt 2, and Opt 3 is based on measurement of Set A.*
* Note: Purpose, such as above “For NW-sided model”, will not be specified in RAN 1 specifications

##### **(High)Proposal 3.1A:**

For NW-sided model, the “beam related information” in a beam report in L1 signaling, at last for BM-Case 1 and for one time instance of BM-Case 2, support the following options:

* Opt 1(w omission): L1-RSRPs and corresponding beam information of Top M beam(s) of a resource set
  + Alt 1: Top M beam(s) is the beams with largest M measured values of L1-RSRPs, where M is configured by gNB
  + FFS: Alt 2: M beams within X dB gap to the largest measured value of L1-RSRP
  + FFS on the maximum value of M (where M >4)
* Opt 2 (w/o omission): All L1-RSRPs of a resource set
  + FFS: without beam information or with best beam index (for differential L1-RSRP reporting, if supported))
* FFS
  + Opt 3: Beam information only of Top M beam(s) of a resource set
  + Opt 4: Opt 3 for one resource set, and Opt 1 or Opt 2 for another resource set.
* FFS on beam information
* Note: Purpose, such as above “For NW-sided model”, will not be specified in RAN 1 specifications

|  |  |
| --- | --- |
| Company | Comments |
| FL | Please read FL’s comments inserted above the proposals first, then provide your views on the Proposal 3.1A. |
| HwHiSi | Support Opt 1.  For Option 2 on the FFS. For the first part of the FFS can it be clarified what the gain would be with this approach? If without beam information, then absolute L1-RSRPs need to be reported. This is more overhead consuming. We think that reporting different L1-RSRP and the indication of the strongest beam would be sufficient for this option.   * Opt 2 (w/o omission): All L1-RSRPs of a resource set   + FFS: ~~without beam information or~~ with best beam index (for differential L1-RSRP reporting, if supported))   For Option 4: Is not needed since separate CSI reports can be configured.  ~~Opt 4: Opt 3 for one resource set, and Opt 1 or Opt 2 for another resource set.~~ |
| TCL | We support Opt 1, either threshold based or number based is reasonable, represents different tradeoffs on configuration flexibility and the performance.  We do not support Opt 2, since weak beams with small RSRP may be overwhelmed by the noise, and it is low quality data for model inference.  For NW-sided model both the beam information and the RSRP is necessary, so we do not support Opt 3 and 4. |
| vivo | If only part of configured beams from UE are reported (e.g., Opt 1), whether the selected beams by UE belongs to the supported pattern of network-sided model is unclear(e.g.,is too large so that supporting all the patterns are unrealistic) considering the pattern misalignment issue. Therefore, option 5 needs to be supported   * Opt 5(revised): Index of a group of beams (identified as subset resource set of a resource set) and all L1-RSRPs of the group of beams. |
| MediaTek | 1. Support both Opt1 and Opt2 (including FFS in Opt2). Opt2 can be used without any beam information. Note that this proposal includes reporting for “one time instance of BM-Case 2”. Therefore, the differential L1-RSRP can be reported when it is defined as the differential between the L1-RSRP (absolute L1-RSRP) of first time instance of BM-Case2 and the rest of time instances of BM Case2.  2. Support Opt3, this proposal is not limited to inference/training/monitoring, for training and monitoring, UE may need to just report beam information (resource ID) for a resource set as ground truth. |
| NTT DOCOMO | Reporting all L1-RSRP of a group of beams is missing in the current proposal. This is useful to enable variable Set B within pre-defined sets. Hence, we suggest the following modification.  For NW-sided model, the “beam related information” in a beam report in L1 signaling, at ~~last~~ least for BM-Case 1 and for one time instance of BM-Case 2, support the following options:   * Opt 1(w omission): L1-RSRPs and corresponding beam information of Top M beam(s) of a resource set   + Alt 1: Top M beam(s) is the beams with largest M measured values of L1-RSRPs, where M is configured by gNB   + FFS: Alt 2: M beams within X dB gap to the largest measured value of L1-RSRP   + FFS on the maximum value of M (where M >4) * Opt 2 (w/o omission): All L1-RSRPs of ~~a resource set~~ a group of resources   + FFS: without beam information or with best beam index (for differential L1-RSRP reporting, if supported))   FFS: a group of resources (e.g., resource set, sub-config within a resource set) |
| QC | This proposal is a follow-up on the following agreement from RAN1 #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**   Which follows up on the “FFS on the report content”. So, a note is added to the updated version in the following to clarify the intent. For other purposes, we can have separate proposals.  Even though Alt. 2 was proposed by some companies, we did not evaluate Alt. 2 in Rel. 18 to assess the benefits of such approach.  Updated Proposal 3.1A  For ~~NW-sided model,~~ the “beam related information” in a beam report in L1 signaling, at last for BM-Case 1 and for one time instance of BM-Case 2, support the following options:   * Opt 1(w omission): L1-RSRPs and corresponding beam information of Top M beam(s) of a resource set   + Alt 1: Top M beam(s) is the beams with largest M measured values of L1-RSRPs, where M is configured by gNB   + ~~FFS: Alt 2: M beams within X dB gap to the largest measured value of L1-RSRP~~   + FFS on the maximum value of M (where M >4) * Opt 2 (w/o omission): All L1-RSRPs of a resource set   + FFS: without beam information or with best beam index (for differential L1-RSRP reporting, if supported)) * ~~FFS~~    + ~~Opt 3: Beam information only of Top M beam(s) of a resource set~~   + ~~Opt 4: Opt 3 for one resource set, and Opt 1 or Opt 2 for another resource set.~~ * FFS on beam information * ~~Note: Purpose, such as above “For NW-sided model”, will not be specified in RAN 1 specifications~~ * Note: the proposal is applicable to the beam report to enable inference of NW-side model. |
| CATT | Add Opt x: the combination of Opt 1/Opt 2 and Opt 3.  For example, when set B is subset pf set A, for training, the UE can report the L1-RSRP of Set B and beam information of Top-M beam of Set A in one report. |
| ZTE | For the second FFS in Opt 1, the number of reported beams can be larger than 4, but it doesn't mean that the configurable M value or the M number of beams within X dB has always to be larger than 4.  For the FFS in Opt 2, the meaning of 'best' should be further clarified. Besides, differential L1-RSRP reporting has already been supported by legacy, and thus the FFS part should be whether to support non-differential L1-RSRP reporting.  Additionally, Opt 3 should be supported for the training/monitoring of classification model at the NW side, where the model label is beam ID of Top-1/K beam in Set A according to Rel-18 evaluations. Accordingly, we have the following revisions.   * Opt 1(w omission): L1-RSRPs and corresponding beam information of Top M beam(s) of a resource set   + Alt 1: Top M beam(s) is the beams with largest M measured values of L1-RSRPs, where M is configured by gNB   + FFS: Alt 2: M beams within X dB gap to the largest measured value of L1-RSRP   + FFS on the maximum value of M (where M ~~>~~ can be larger than 4) * Opt 2 (w/o omission): All L1-RSRPs of a resource set   + FFS: without beam information (for non-differential L1-RSRP reporting, if supported) or with best beam index associated with the largest measured value of L1-RSRP ~~(for differential L1-RSRP reporting, if supported)~~ * ~~FFS~~    + Opt 3: Beam information only of Top M beam(s) of a resource set   + Opt 4: Opt 3 for one resource set, and Opt 1 or Opt 2 for another resource set. |
| Panasonic | Support Option 1. |
| Xiaomi | Support opt 1 and opt 2 at least.  While for Opt 3, it can be used for performance monitoring to report the measured beam information of Top-K beams. Opt 4 can be discussed after Opt 3. |
| NEC | Support using omission to reduce overhead.  We only need to omit those very weak beams which cannot contribute to model inference. With that said, we support alternatives to report beams higher than a threshold value, or we could be fine to Alt 2 with the following change  Alt 2: ~~M~~all beams within X dB gap to the largest measured value of L1-RSRP  On the other hand, if no change of the Alt2, we need to understand how “M” is determined, it may be determined by the UE based on a minimal value that NW configures. |
| New H3C | OK |
| InterDigital | Support both options. |
| Ericsson | Support option 1.  Option 2. We agree on the FFS intention on how to save reporting overhead. However, for now the FFS is unclear, prefer to keep it open for now.   * Opt 2 (w/o omission): All L1-RSRPs of a resource set   FFS: reporting format details ~~without beam information or with best beam index (for differential L1-RSRP reporting, if supported))~~ |
| KDDI | Support both options. |
| SPRD | Support option1 and option2. |
| LG | Support Opt 1. |
| Fujitsu | We suggest to have separate proposal for inference, monitoring and training data collection, since the different option may be suitable for different purpose.  One example is for monitoring, if the performance metric is beam prediction accuracy, then RSRP information may not be needed for the reporting. |
| Google | Is it correct that this is for model inference? For model inference, we do not know how option 1 can work. Besides, model inference does not require UE to report too many L1-RSRPs, why is the overhead problematic? |
| CMCC | Support option 1 for training, option 1 or 2 for inference, option 4 for monitoring. |
| Apple | The FFS before “option 2” under alt. 1 should be removed. In the Rel-18 study, at least we evaluated RSRPs within a range towards the RSRP of the strongest beam.   * Opt 1(w omission): L1-RSRPs and corresponding beam information of Top M beam(s) of a resource set   + Alt 1: Top M beam(s) is the beams with largest M measured values of L1-RSRPs, where M is configured by gNB   + ~~FFS:~~ Alt 2: M beams within X dB gap to the largest measured value of L1-RSRP   + FFS on the maximum value of M (where M >4) |
| CAICT | General fine with the proposal. For NW-sided inference, L1-RSRP should at least be included and Opt 1 should be baseline. |
| Lenovo | Support option 1 with Alt 2 and option 2. |
| Fraunhofer | Support Option 1 and Option 2. |
| OPPO | Support Option 2 and open to discuss more details on Option 1. |

#### Issue #2: Beam information in L1 report

At least for NW-side model, for the reported beam information

* For content Opt 1(w omission),
  + Opt 0: legacy CRI/SSBRI, (i.e., index of resource in a resource set)
  + *FL’s comments: No problem to support as default*
  + ~~Opt 1: beam indexes are reported based on a bitmap, where bitmap indicating RS index of a resource set.~~
  + *FL’s comments: Only when all absolute L1-RSRPs. Cannot support differential L1-RSRP*
  + *From FL point of view, differential L1-RSRP can save more overhead. therefore, no need to consider this.*
  + FFS: Opt 2: The beam index with largest measured value of L1-RSRP, and a bitmap, where bitmap indicating RS index of a resource set
  + *FL’s comments: Only when differential L1-RSRP is supported.*
  + *In additional, the omission is based on L1-RSRSP, therefore, it cannot support Set based beam ID.*
* For content Opt 2 (w/o omission):
  + ~~Opt 2: No beam index.~~
  + *FL’s comments: This is possible when all absolute L1-RSRPs. Otherwise, one beam ID is needed. but no motivation to support absolute L1-RSRPs*
  + Opt 3: Only one beam index with largest measured value of L1-RSRP (i.e., CRI/SSBRI in a resource set)
  + *FL’s comments: This is useful if differential L1-RSRSP is supported.*
* For Opt 3 (only beam information): Beam information of Top M beam(s) of a resource set, support
  + Opt A: legacy CRI/SSBRI, (i.e., index of resource in a resource set)
    - Order CIR/SSBRI according to the values of L1-RSRPs of Top M beam(s)
  + *FL’s comments: Order of Top M is important therefore, Opt A needed. .*
* For content Opt 4 (one report associated with two resource sets):
  + Opt 5: Index of a group of beams (identified as subset resource set of a resource set)
  + *FL’s comments: Index can be used to indicate one resource set, or from predefined/configured multiple sets*

##### **(High)Proposal 3.2A:**

For NW-sided model, “beam information” in a beam report at least for L1 signaling, at last for BM-Case 1 and for one time instance of BM-Case 2,

* For Opt 1(w omission): L1-RSRPs and corresponding beam information of Top M beam(s) of a resource set, support:
  + Opt A: CRI/SSBRI, (i.e., index of resource in a resource set)
  + FFS: Opt B: The beam index with largest measured value of L1-RSRP, and a bitmap, where bitmap indicating RS index of a resource set, if differential L1-RSRSP is supported.
* For content Opt 2 (w/o omission): all L1-RSRPs of a resource set, support
  + Opt D: Only one beam index with largest measured value of L1-RSRP (i.e., CRI/SSBRI in a resource set), if differential L1-RSRSP is supported.
* ~~FFS:~~ For Opt 3 (only beam information): Beam information only of Top M beam(s) of a resource set, support
  + Opt A: CRI/SSBRI, (i.e., index of resource in a resource set)
    - The ranking of Top M beam(s) is conveyed by the order of CRI/SSBRI
* ~~FFS:~~ For content Opt 4 (one report associated with two resource sets):
  + FFS on details
* [Note: the content options are separated discussed]

|  |  |
| --- | --- |
| Company | Comments |
| FL | Please read FL’s comments inserted above the proposals first, then provide your views on the Proposal 3.2A.  My intention is to resolve the content in proposal 3.1A, this doesn’t mean Opt 3/4 are supported. |
| HW/HiSi | For Opt 1 support in principle. For the FFS we have a comment. Depending in the size of the measured set and the number of beams to be reported, either the legacy CRI approach or a bitmap has less overhead. It should therefore be allowed to let the UE flexibly determine depending on the configuration, which format to use for reporting.  For Opt 4, can the intention be clarified? Does it also cover the possibility t measure a larger Set A? We think it is an important issue to consider and was also part of the FL summary from RAN1#116b. Since in this case it could be more than 2 resource sets, we suggest to change “two” to “multiple”  **Suggested udpates**:  For NW-sided model, “beam information” in a beam report at least for L1 signaling, at last for BM-Case 1 and for one time instance of BM-Case 2,   * For Opt 1(w omission): L1-RSRPs and corresponding beam information of Top M beam(s) of a resource set, support:   + Opt A: CRI/SSBRI, (i.e., index of resource in a resource set)   + FFS: Opt B: The beam index with largest measured value of L1-RSRP, and a bitmap, where bitmap indicating RS index of a resource set, if differential L1-RSRSP is supported. The UE determines whether to report OptA or OptB depending on the configuration of the size of the measured and the number of M. * For content Opt 2 (w/o omission): all L1-RSRPs of a resource set, support   + Opt D: Only one beam index with largest measured value of L1-RSRP (i.e., CRI/SSBRI in a resource set), if differential L1-RSRSP is supported. * ~~FFS:~~ For Opt 3 (only beam information): Beam information only of Top M beam(s) of a resource set, support   + Opt A: CRI/SSBRI, (i.e., index of resource in a resource set)     - The ranking of Top M beam(s) is conveyed by the order of CRI/SSBRI * FFS: For content Opt 4 (one report associated with multiple ~~two~~ resource sets):   + FFS on details     - [Note: the content options are separated discussed] |
| TCL | We support Opt 1, either threshold based or number based is reasonable, represents different tradeoffs on configuration flexibility and the performance.  We do not support Opt 2, since weak beams with small RSRP may be overwhelmed by the noise, and it is low quality data for model inference.  Opt 3 is OK. The benefit of Opt 4 is not clear. |
| vivo | Option 5 needs to be considered as a candidate content, and it can be used for content Opt 1/2/3 as well. We, thus, propose,  For NW-sided model, “beam information” in a beam report at least for L1 signaling, at last for BM-Case 1 and for one time instance of BM-Case 2,   * For Opt 1(w omission): L1-RSRPs and corresponding beam information of Top M beam(s) of a resource set or a subset resource set, support:   + Opt A: CRI/SSBRI, (i.e., index of resource in a resource set)   + FFS: Opt B: The beam index with largest measured value of L1-RSRP, and a bitmap, where bitmap indicating RS index of a resource set, if differential L1-RSRSP is supported. * For content Opt 2 (w/o omission): all L1-RSRPs of a resource set or a subset resource set, support   + Opt D: Only one beam index with largest measured value of L1-RSRP (i.e., CRI/SSBRI in a resource set), if differential L1-RSRSP is supported. * ~~FFS:~~ For Opt 3 (only beam information): Beam information only of Top M beam(s) of a resource set or a subset resource set, support   + Opt A: CRI/SSBRI, (i.e., index of resource in a resource set)     - The ranking of Top M beam(s) is conveyed by the order of CRI/SSBRI * ~~FFS:~~ For content Opt 4 (one report associated with two resource sets):   + FFS on details * For content Opt 5:Index of a group of beams can be identified as subset resource set of a resource set * [Note: the content options are separated discussed] |
| QC | We understand FL’s intention, and we believe it would be better to fist converge on 3.1, and then discuss the related options for beam information. |
| CATT | For Opt B, the beam index with largest measured value of L1-RSRP refers to the beam index with largest measured value of L1-RSRP within a bitmap. For example, if the size of bitmap is 8 bit, then the beam index with largest measured value of L1-RSRP can be represent by a 3-bit beam index. |
| ZTE | For Opt 1, if the number of beams to be reported is configured by the NW (e.g., 1/2 or 1/4 of beams in Set C according to Rel-18 evaluations), the overhead for reporting legacy CRI/SSBRI increase linearly with the number of beams to be reported and becomes considerably huge. In this case, bitmap should be used instead for reporting overhead reduction without compromising any performance. Bitmap should always be reported, and whether to report the beam index with largest measured value of L1-RSRP depends on whether non-differential L1-RSRSP is supported.  Additionally, for the differential L1-RSRSP part, we have the same comments as Proposal 3.1A. Therefore, we suggest the following revisions.   * For Opt 1(w omission): L1-RSRPs and corresponding beam information of Top M beam(s) of a resource set, support:   + Opt A: CRI/SSBRI, (i.e., index of resource in a resource set)   + ~~FFS:~~ Opt B: ~~The beam index with largest measured value of L1-RSRP, and~~ a bitmap, where bitmap indicating RS index of a resource set, with or without the beam index with largest measured value of L1-RSRP depending on whether non-differential L1-RSRSP is supported. * For content Opt 2 (w/o omission): all L1-RSRPs of a resource set, support   + Opt D: Only one beam index with largest measured value of L1-RSRP (i.e., CRI/SSBRI in a resource set)~~, if differential L1-RSRSP is supported.~~ |
| Panasonic | We are not sure about difference between Proposal 3.1 A and Proposal 3.2A. The difference is “related” wording? Can FL clarify it?   * Proposal 3.1A: “For NW-sided model, the “beam related information” in a beam report in L1 signaling, at last for BM-Case 1 and for one time instance of BM-Case 2 ….” * Proposal 3.2A: “For NW-sided model, “beam information” in a beam report at least for L1 signaling, at last for BM-Case 1 and for one time instance of BM-Case 2….”   @FL, if these 2 proposals are saying the same thing, should we discus one of them only? |
| Xiaomi | Support Opt A in Opt 1. Whether to support Opt B depends on the number of resource in the resource set and the value of M.  For Opt D in Opt 2, for BM Case 2, it is possible that only the beam index with largest measured value of L1-RSRP in the first time instance is needed. But in the later time instance, the differential L1-RSRP can be repirted refer to the absolute RSRP of the first time instance. So we suggest to add a note after Opt D.   * For content Opt 2 (w/o omission): all L1-RSRPs of a resource set, support   + Opt D: Only one beam index with largest measured value of L1-RSRP (i.e., CRI/SSBRI in a resource set), if differential L1-RSRSP is supported.     - Note: at least for BM case 1 and for at least one time instance for BM case 2. |
| NEC | We could be fine to support opt 1 opt B. |
| New H3C | OK |
| InterDigital | We do not support Opt 3 and Opt 4. For those options, further discussion is needed. |
| Ericsson | First, we should agree on **Proposal 3.1A.**  Furthermore. Unclear why we need to discuss best beam index. The following should be more clear.  ……..   * For Opt 1(w omission): L1-RSRPs and corresponding beam information of Top M beam(s) of a resource set, support:   + Opt A: CRI/SSBRI, (i.e., index of resource in a resource set)   + FFS: Opt B: The ~~beam~~ RS index for the ~~with~~ largest measured ~~value of~~ L1-RSRP of a resource set, and a bitmap, where bitmap indicating RS index of a resource set, if differential L1-RSRSP is supported. * For content Opt 2 (w/o omission): all L1-RSRPs of a resource set, support   + Opt D: Only one ~~beam~~ RS index for the ~~with~~ largest measured value of L1-RSRP (i.e., CRI/SSBRI in a resource set), if differential L1-RSRSP is supported. |
| LG | Support Option 1-A. I think differential RSRP can be considered in this option.  Question to FL: is it the intention that support multiple options in this proposal? |
| Fujitsu | We also suggest to have separate proposal for inference, monitoring and training data collection, since the different option may be suitable for different purpose.  With the current formulation, does it mean training data collection will be via L1 singling? |
| Google | We are a bit uncertain for the use case of w/ omission. |
| CMCC | Support Opt 1 opt B. Similar comment as HW, the overhead of legacy CRI approach and bitmap changes with different size of the measured set and the number of beams to be reported. A pre-defined rule can be used to determine the format of reporting. |
| Apple | We support Option 1 Option B, and suggest the removal of the FFS to be on the same footing as other proposals.   * + ~~FFS:~~ Opt B: The beam index with largest measured value of L1-RSRP, and a bitmap, where bitmap indicating RS index of a resource set, if differential L1-RSRSP is supported. |
| CAICT | Fine with FL’s proposal and Opt. 1 and 2 is preferred. |
| Lenovo | We think CRI/SSBRI based beam report should be the baseline for option 1. |
| Fraunhofer | Support Option 1 and Option 2. |
| OPPO | In our reading, this proposal highly relates to Proposal 3.1A. It may sound better to discuss the reporting contents and clarifying the beam information together. |

#### Issue #3: Quantization for beam report

##### **(High)Proposal 3.3A:**

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

* Option 1: Support differential L1-RSRP reporting with legacy quantization step and range
* Option 2: Support differential L1-RSRP reporting with larger quantization step(s) than legacy
  + FFS: with smaller range(s) for differential L1-RSRP than legacy
  + FFS: step size(s) for absolute L1-RSRP, step size(s) for differential L1-RSRP, range(s) for differential L1-RSRP
* Option 1 or Option 2 is configured by gNB

|  |  |
| --- | --- |
| Company | Comments |
| FL | FL’s comments:   * Delete: FFS on whether to support absolute L1-RSRP reporting (for all beams in a set)   + According to Fraunhofer’s calculation (Thank to Fraunhofer), the overhead is high with absolute L1-RSRP reporting, even when total number of beam is small.   + In current beam report, even with M=4 beams, we support differential L1-RSRP report * Delete: FFS on whether to support reporting the normalized L1-RSRP measurement instead of actual L1-RSRP values   + The saving is just L1-RSRP of absolute (which is normalized to 1). No much needed. * Reason to support larger step size: majority companies support, evaluations in SI.   + One company support finer step, which didn’t get support in SI. * Reason to FFS smaller range:   + Supported by the evaluation in SI.   + No much explicitly mentioned in companies view.   + This may be related to omission * One more step on configurable. I guess no need to limit this to special case. And shall be controlled by gNB * Details for high layer can be further discussed. |
| HW/HiSi | In our understanding Option 1 is already supported and does not need an extra agreement, or? In this case, we may not need a proposal for Opt1? Whether the quantization size needs to be changed also depends on other discussions on overhead saving.  Suggested update:  At least for NW sided model, for the quantization of a reported L1-RSRP value at least for the report in L1 signaling, study   * ~~Option 1: upport differential L1-RSRP reporting with legacy quantization step and range~~ * ~~Option 2:~~ whether to additionally Support differential L1-RSRP reporting with larger quantization step(s) than the already supported legacy quantization step and range   + FFS: with smaller range(s) for differential L1-RSRP than legacy   + FFS: step size(s) for absolute L1-RSRP, step size(s) for differential L1-RSRP, range(s) for differential L1-RSRP * ~~Option 1 or Option 2 is configured by gNB~~ |
| TCL | We suggest to change the Option 2 as follows and add one FFS.   * Option 2: Support differential L1-RSRP reporting with ~~larger~~ different quantization step(s) than legacy   + FFS: the quantization steps are uniform or variable.   + FFS: with smaller range(s) for differential L1-RSRP than legacy   + FFS: step size(s) for absolute L1-RSRP, step size(s) for differential L1-RSRP, range(s) for differential L1-RSRP |
| MediaTek | We support to add normalized L1-RSRP as one of the options. The benefit of normalized L1-RSRP is not in bit saving, but to more correctly report for the model input. Quantizing a L1-RSRP will lose some information from the actual value. For differential L1-RSRP approach, NW will normalize two quantized L1-RSRP value, more information will be lost when normalizing on two quantized values; while for normalized L1-RSRP approach, UE can normalize two ACTUAL L1-RSRP values and directly quantize the result (with same quantization step as normalized L1-RSRP). Only one value is quantized, so the information loss for model input should be lesser. Note that NW can still obtain the L1-RSRP for each Set B beam if NW signals the mean/variance (or max/min) to UE to do normalization. |
| NTT DOCOMO | Support. |
| CATT | We are ok with the proposal |
| ZTE | General fine. Per our understanding, even if larger quantization step size and smaller quantization range are both supported, they don't need to be bundled together. For example, it may be possible that the differential L1-RSRP reporting is enabled with larger quantization step size and legacy quantization range, OR with legacy quantization step size and smaller quantization range. Therefore, we can separately discuss the two issues as follows.   * Option 1: Support differential L1-RSRP reporting with legacy quantization step and range * Option 2: Support differential L1-RSRP reporting with larger quantization step(s) than legacy   + ~~FFS: with smaller range(s) for differential L1-RSRP than legacy~~   + FFS: step size(s) for absolute L1-RSRP, step size(s) for differential L1-RSRP, ~~range(s) for differential L1-RSRP~~ * FFS Option 3: Support differential L1-RSRP reporting with smaller range(s) for differential L1-RSRP than legacy * Option 1 or Option 2 is configured by gNB |
| Panasonic | As mentioned by other company, Option 1 is as same as legacy one.  We are fine with updates from HW. |
| Xiaomi | For Option 2, in my understanding, since the number of reported beams will be more than 4, the range of differential L1-RSRP need to be increased. So we suggest to revise Option 2 as below:   * Option 2: Support differential L1-RSRP reporting with larger ranges/quantization step(s) than legacy   In addition, what is the motivation of the FFS? We suggest to delete it. |
| NEC | Support option 2, and especially for BM case 2 we support to report RSRP/differential RSRP with a larger step size (less bits) for those historical measurements obtained earlier. |
| New H3C | OK |
| InterDigital | We support Option 1.We do not support Option 2. |
| Ericsson | Support update by HW. |
| KDDI | Support. |
| SPRD | OK with this proposal. |
| LG | We support option 1. |
| Fujitsu | We think another option is to use smaller quantization steps or non-differential L1-RSRP. We suggest the following update.  *At least for NW sided model, for the quantization of a reported L1-RSRP value at least for the report in L1 signaling, ~~support~~ further discuss*   * *Option 1: Support differential L1-RSRP reporting with legacy quantization step and range* * *Option 2: Support differential L1-RSRP reporting with larger quantization step(s) than legacy*   + *FFS: with smaller range(s) for differential L1-RSRP than legacy*   + *FFS: step size(s) for absolute L1-RSRP, step size(s) for differential L1-RSRP, range(s) for differential L1-RSRP* * *Option 3: Support differential L1-RSRP reporting with smaller quantization step(s) than legacy* * *Option 4: Support non-differential L1-RSRP reporting* * *~~Option 1 or Option 2 is~~ ~~configured by gNB~~* |
| Google | Agree with the update from Fujitsu |
| CMCC | Ok. |
| Apple | We support Option 1, but if Option 2 is listed as an option, it would be fine if Proposals 3.1A and 3.2A options are treated in a similar fashion. |
| CAICT | Support HW’s proposal. |
| Lenovo | Option 1 is enough. |
| Fraunhofer | Agree. |
| OPPO | We think legacy quantization works good, and we failed to find strong motivation to specify additional quantization steps and/or ranges. |

#### Issue #4: Content for data collection for training in higher layer

##### **(low)Potential Proposal 3.4A: (for comment collection)**

For NW-sided model, the content for data collection for training in high layer, at last for BM-Case 1 and for one time instance of BM-Case 2, FFS the following options:

* Opt 1(w omission): L1-RSRPs and corresponding beam information of Top M beam(s)
  + FFS on the maximum value of M and how to determinate M, FFS:
    - Alt 1: M strongest with highest L1-RSRP, where M is configured by gNB
    - Alt 2: M beams within X dB gap to the highest L1-RSRP
  + FFS on the beam information
* Opt 2 (w/o omission): All L1-RSRPs
  + FFS: without beam information or with best beam index (for differential L1-RSRP reporting, if supported))
* Opt 3: Beam information of Top M beam(s)
* Opt 4: Beam information of Top M beam(s) of one set, and Opt 1 or Opt 2 for the second set.
* FFS on the one or more than one resource set associated with one high layer report.

|  |  |
| --- | --- |
| Company | Comments |
| FL | RAN 2 already agreed to use MDT.  Slightly modify based on proposal 3.1A.  More motivation to support opt 3 and opt 4 .  No need to limited to one resource set |
| HW/HiSi | First priority should be given to proposal 3.1 and 3.2, L1-signaling. The higher layer signaling proposal could be deprioritized from the RAN1 perspective.  It seems we are discussion here the contents of the signaling and (Similar to 3.1A) and they are the same. In the case, if the proposal should be treated, the “higher layer” could be removed from the proposal:  A potential update of the proposal is given below. **(low)Potential Proposal 3.4A: (for comment collection)** For NW-sided model, the content for data collection for training ~~in high layer~~, at last for BM-Case 1 and for one time instance of BM-Case 2, FFS the following options:   * Opt 1(w omission): L1-RSRPs and corresponding beam information of Top M beam(s)   + FFS on the maximum value of M and how to determinate M, FFS:     - Alt 1: M strongest with highest L1-RSRP, where M is configured by gNB     - Alt 2: M beams within X dB gap to the highest L1-RSRP   + FFS on the beam information * Opt 2 (w/o omission): All L1-RSRPs   + FFS: without beam information or with best beam index (for differential L1-RSRP reporting, if supported)) * Opt 3: Beam information of Top M beam(s) * ~~Opt 4: Beam information of Top M beam(s) of one set, and Opt 1 or Opt 2 for the second set.~~ |
| TCL | We support Opt 1, either threshold based or number based is reasonable, represents different tradeoffs on configuration flexibility and the performance.  We do not support Opt 2, since weak beams with small RSRP may be overwhelmed by the noise, and it is low quality data for model inference.  Opt 3 is OK. The benefit of Opt 4 is not clear. |
| vivo | OK |
| CATT | This proposal can be used for L1 signaling as well. |
| Fujitsu | Is the proposal for the reporting of Set B beams or Set A beams? This should be clarified. |
| Google | For data collection, we think w/o omission should be considered. |

#### Issue #5: Whether to support multiple time instances in one report

Companies support it:

* (12 +2) Samsung, vivo, Interdigital, Google, Sharp, MTK, CATT, Nokia, KDDI, ZTE, xiaomi, Ericsson?
* Spreadtrum: Yes for training
* LGE: yes for high layer

Companies not support it:

* (1) Huawei/HiSi

##### **(low)Potential Proposal 3.5A: (for BM-Case 2)**

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

* FFS on the method to reduce overhead, including differential L1-RSRP reporting over multiple time instances
* FFS on how to indicate the time stamp information
* Note: Purpose, such as above “For NW-sided model, for BM-Case 2”, will not be specified in RAN 1 specifications

|  |  |
| --- | --- |
| Company | Comments |
| FL | Collect views in 1st round.  No need to limit the purpose of this report. Companies are invited to provide use cases. |
| HW/HiSi | This is not a critical issue and can be supported with legacy.  Additionally, before discussing the report of multiple past time instances, we should discuss the measurements. This discussion would be more important, since before reporting we need to have something to measure. Also, the measurement could apply to both BM Case 1 and BM Case 2. We are therefore making the following updated generic proposal: **Updated (low)Potential Proposal 3.5A: (for BM-Case 2)** For NW-sided model, ~~for BM-Case 2~~, support ~~to report of~~ measurement ~~results~~ of multiple resource sets ~~past time instances~~ in one CMR procedure ~~report.~~   * FFS on the method to reduce overhead, including differential L1-RSRP reporting over multiple time instances * FFS on how to indicate the time stamp information   Note: Purpose, such as above “For NW-sided model, ~~for BM-Case 2~~”, will not be specified in RAN 1 specifications |
| TCL | Support differential L1-RSRP over multiple time instances and the reference RSRP may be chosen by position or value.  The time stamp information can also be compressed by reporting a reference time and the time offset. |
| vivo | General support the proposal. Further, we support time domain compression of beam indication to further reduce report overhead within a report. Thus,  For NW-sided model, for BM-Case 2, support to report of measurement results of multiple past time instances in one report.   * FFS on the method to reduce overhead, ~~including~~, e.g., differential L1-RSRP reporting over multiple time instances, and time domain compression of beam indication * FFS on how to indicate the time stamp information   Note: Purpose, such as above “For NW-sided model, for BM-Case 2”, will not be specified in RAN 1 specifications |
| CATT | OK with the proposal. |
| ZTE | Support vivo’s revision. |
| SPRD | We did not see the benefit of carrying multiple historical time instances measurement results in one report. |
| LG | Support the FL proposal. |
| Fujitsu | What’s the benefit to report results of multiple past time instances? |
| Google | Support |
| CMCC | Support vivo’s revision. |

#### Issue #6: Max number of reported beam related information in one report in L1 signaling

##### **Questions to answer**

A: Max number in a report for BM case 1

B: Max number in a report if BM-Case 2 is supported

|  |  |
| --- | --- |
| Company | Comments |
| FL | Collect views in 1st round.  Note: in the end, it will be configured based on UE capability. |
| HW/HiSi | A: For BM Case 1 Max number could be at least 32 beams.  B: For BM Case 2 we anyway need to increase the number of beams in a report, the UE capability of UE for BM Case 1 would not be increased beyond BM case 2. For Case 2, we could have 32 beams in total. |
| TCL | A: 256  B: 256 |
| Ericsson | A: 256  B: 256 |
| LG | A: 32  B: 128 |
| Google | If we are considering the UE complexity, we think a more critical issue is the maximum number of beams to be measured.  For A and B, it depends on the use case, and whether it is for UE side model or NW side model. |

### 3.5 2st Round discussion

#### Issue #1: L1 Report content for NW-sided model

Way to discussion combined:

**Proposal 3.1B**

For NW-sided model, the “beam related information” in a beam report in L1 signaling, at least for BM-Case 1 and for one time instance of BM-Case 2, support the following options:

* Opt 1(w omission): L1-RSRPs and corresponding beam information of Top M beam(s) of a resource set
  + FFS
    - 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
  + FFS on the maximum value of M (where M can be larger than 4)
  + FFS on beam information
  + *For information: this can be used for inference/monitoring/(cannot work for training by itself)*
* Opt 2 (w/o omission): All L1-RSRPs of a resource set
  + with beam index (i.e., CRI/SSBRI) with largest measured values of L1-RSRPs (for differential L1-RSRP reporting)
  + *For information: this can be used for inference/monitoring/training*
* Opt 3: Only beam index (i.e., CRI/SSBRI) of Top M beam(s) of a resource set
  + *For information: this can be used for monitoring/training (cannot work for training by itself)*
* FFS: the combination of Opt 3 (Beam index (i.e., CRI/SSBRI)), and Opt 1 or Opt 2 (L1-RSRP and beam index (i.e., CRI/SSBRI))
  + *For information: this can be used for monitoring/training*
* FFS: whether a resource set can be a subset of CMR and details of beam information for each Opt
* Note: Purpose, such as above “For NW-sided model”, will not be specified in RAN 1 specifications

**Proposal 3.1B**

For NW-sided model, at least for inference the “beam related information” in a beam report in L1 signaling, at least for BM-Case 1 and for one time instance of BM-Case 2, support the following options:

* Opt 1(w omission): L1-RSRPs and corresponding beam information of Top M beam(s) of a resource set
  + FFS
    - 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
  + FFS on the maximum value of M (where M can be larger than 4)
  + FFS on beam information
* Opt 2 (w/o omission): All L1-RSRPs of a resource set
  + with beam index (i.e., CRI/SSBRI) with largest measured values of L1-RSRPs (for differential L1-RSRP reporting)
* FFS: whether a resource set can be a subset of CMR and details of beam information for each Opt
* Note: Purpose, such as above “For NW-sided model, at least for inference”, will not be specified in RAN 1 specifications

For NW-sided model, at least for monitoring,in a beam report initiated by network, based on one [or two] measurement resource set, support the report of more than 4 beam related information in L1 signaling,

* wherethe “beam related information”, at least for BM-Case 1 and for one time instance of BM-Case 2, support the following options:
* Opt 1(w omission): L1-RSRPs and corresponding beam information of Top M beam(s) of a resource set
  + FFS
    - 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
  + FFS on the maximum value of M (where M can be larger than 4)
  + FFS on beam information
* Opt 2 (w/o omission): All L1-RSRPs of a resource set
  + with beam index (i.e., CRI/SSBRI) with largest measured values of L1-RSRPs (for differential L1-RSRP reporting)
* Opt 3: Only beam index (i.e., CRI/SSBRI) of Top M beam(s) of one resource set
* FFS: Opt 4: the combination of Opt 3 (Beam index (i.e., CRI/SSBRI)), and Opt 1 or Opt 2 (L1-RSRP and beam index (i.e., CRI/SSBRI))
  + FFS based on one or two measurements set
* ~~FFS: whether a resource set can be a subset of CMR and details of beam information for each option~~
* Note: Purpose, such as above “For NW-sided model, at least for monitoring”, will not be specified in RAN 1 specifications

|  |  |  |
| --- | --- | --- |
| Company | Prefer  A or B | Comments |
| FL | A | Pls indicate which on do you prefer. From FL point of view, I prefer to not mention the purpose but directly discuss about the report content  Secondly, pls indicate your comments of each option  FFS on “subset of CMR” is from DoCoMo, who thinks the UE can measures multiple set, but the only choose one subset of resources to report. E.g., assuming NW will support variable Set B.   * Please also indicate whether you support it or not.   For opt 3, I think it is usually for monitoring. for Opt 4, as well. but opt 4 can be configured with different report. Well, similar as the discussion online today, I am not sure whether NW can manage the link between Set B measurement and Set A (e.g., best beam). Or even for Opt 1/2, whether we need to configure two CMRs. |
| OPPO | B | With the specific purpose, we can understand the intentions of the proposals more clearly.  As for the subset of CMR for tailoring variable Set B, from evaluation results of variable Set B from R18 SI, we failed to find its benefits over fixed Set B. Then we have no strong reason to support it. |
| MediaTek | A | For Opt2, we support to add “without any beam information” back because this proposal includes reporting for “one time instance of BM-Case 2”. The differential L1-RSRP can be reported for all beams without any beam ID when it is defined as the differential between the L1-RSRP (absolute L1-RSRP) of the first time instance of BM-Case2 and the rest of time instances of BM Case2. |
| Hw/HiSi | A | Opt1: Support in principle. But Alt1 should not be FFS since it already is supported in legacy, except that M can be larger than 4, but on this we have already agreed. It would also be good to set a number on M. Since for BM Case 2 Set A can be Set B, we suggest to trigger a discussion to support a relatively larger value for M, e.g. 32   * *Opt 1(w omission): L1-RSRPs and corresponding beam information of Top M beam(s) of a resource set*   + *~~FFS~~*     - *Alt 1: Top M beam(s) is the beams with largest M measured values of L1-RSRPs, where M is configured by gNB*     - *FFS: Alt 2: All beams within X dB gap to the largest measured value of L1-RSRP*   + *FFS on the maximum value of M (where M can be larger than 4, e.g. 32)*   + *FFS on beam information*   Opt 2, Opt 3 ok.  **Not support the FFS to combine of Opt 3, with Opt 1 or Opt 2**. Configuring in separate reports is more efficient, since Opt 1 or Opt 2 are needed for inference anyway, then additionally Opt 3 could be used to separately report the label.  ~~FFS: the combination of Opt 3 (Beam index (i.e., CRI/SSBRI)), and Opt 1 or Opt 2 (L1-RSRP and beam index (i.e., CRI/SSBRI))~~  For the newly added FFS, for the support of a large Set A, we think it also import to study whether multiple resource sets can be part of the CMR procedure and suggest to update accordingly:   * FFS: whether multiple ~~a~~ resource set(s) can be in one ~~a subset of~~ CMR procedure and details of beam information for each Opt |
| Intel | A | To capture the key information from version ‘B’, we could simply add a qualifier for Opt 3:   * Opt 3: Only beam index (i.e., CRI/SSBRI) of Top M beam(s) of a resource set * Applicable at least for monitoring.   We support the suggestion to remove FFS for Alt 1 for Opt. 1.  Also, we prefer to remove Opt. 4. Any potential signalling enhancements can be considered later if justified. It would not be appropriate to list Opt. 3 at the same level as Options 1, 2, 3. |
| Apple | B | * We support 3.1B.   On Opt 1 (copied below), Actually we support Alt. 2, which is better than Alt. 1 in our view. At this time it okay to keep both Alt. 1 and Alt. 2 as FFS.   * On Opt 1(w omission): L1-RSRPs and corresponding beam information of Top M beam(s) of a resource set   + FFS     - 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 |
| ZTE | A | As agreed before that the purpose of UE reporting is transparent to the UE, then we prefer not to mention the purpose in the main bullet and to directly discuss the content.  For the FFS on “subset of CMR”, the method of reporting a pre-defined beam subset would increase the overhead for RS transmission and UE measurement compared with fixed Set B without provide any clear benefit. Additionally, regarding the UE reporting of partial measurement results, Opt 1 is more reasonable as verified in Rel-18.  For opt 4, we are fine to keep it for further study. The sharing of one report setting is beneficial from configuration overhead perspective. On the contrary, if two reporting settings are configured, the time instances of the two reports shall be very close to each other, which may be restrictive in some cases. |
| InterDigital | A | Prefer A as the purpose of UE reporting is not needed for specification support. |
| TCL | A | We support Option A and suggest to add an FFS to capture the differences between training, inference and monitoring.   * FFS: The maximum value of M could be different in training, inference and monitoring. |
| CEWiT | A | We prefer A as the purpose of UE report is not necessary. We support Option 1. |
| CATT | A | For opt 1, we agree with HW that Alt 1 has been supported in legacy L1-RSRP reporting, so the FFS before Alt 1 should be removed.  We are ok with Opt2 and Opt3  For first FFS, we prefer to keep it as a FFS, since it can be used for model training, especial for BM case 1. It can be used to ensure report input and label in same time without additional time correlation indication. Also, it is can work with Alt.2 (one *CSI-ResourceConsigId* is configured for both Set A and SetB) in last meeting agreement. It’s one potential method and should be kept under proposal version A.  For new added FFS, agree with HW’s version：  FFS: whether multiple ~~a~~ resource set(s) can be in one ~~a subset of~~ CMR procedure and details of beam information for each Opt |
| NEC | B | For inference and for monitoring the required report can be different, thus it may be easier if we can discuss separately.  We also support“subset of CMR” is from DoCoMo to configure multiple subsets for Set B, or multiple Set Bs, then, only one subset or only one Set B is reported to NW. |
| QC | B | The content for each purpose is different, so the obvious way forward is to discuss the contents for each purpose **separately**.  The benefits of Opt. 1, Alt. 2 should be justified compared to Opt. 1, Alt. 1. Opt. 1, Alt. 2 leads to variable payload overhead, which would involve a lot of specification effort to enable such a mechanism, whereas Opt. 1, Alt. 1 is based on legacy design with fixed payload overhead. The necessity and potential advantages of Opt. 1, Alt. 2 should be justified.  As a first step, we can address the FFS from the following agreement:  **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**   Whether the purpose is specified or not depends on the “content” of the collected data which may or may not be different for different purposes. Report content may be different for different purposes, and it should be discussed separately.  For this purpose, we should explicitly declare what the context is. So, here’s the updated version:  **Updated Proposal 3.1B**  For NW-sided model, ~~at least~~ for inference the “beam related information” in a beam report in L1 signaling, at least for BM-Case 1 and for one time instance of BM-Case 2, support the following options:   * Opt 1(w omission): L1-RSRPs and corresponding beam information of Top M beam(s) of a resource set   + FFS     - 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~~   + FFS on the maximum value of M (where M can be larger than 4)   + FFS on beam information * Opt 2 (w/o omission): All L1-RSRPs of a resource set   + with beam index (i.e., CRI/SSBRI) ~~with~~ of only the largest measured values of L1-RSRPs (for differential L1-RSRP reporting)   + FFS on the maximum number of L1-RSRPs * FFS: whether a resource set can be a subset of CMR and details of beam information for each Opt   Note: Purpose, such as above “For NW-sided model, ~~at least~~ for inference”, ~~will~~ may not be specified in RAN 1 specifications |

#### Issue #3: Quantization for beam report

##### **(High)Proposal 3.3A:**

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

* Option 1: Support differential L1-RSRP reporting with legacy quantization step and range
* FFS: Option 2: Support differential L1-RSRP reporting with larger quantization step(s) than he already supported legacy quantization step and range
  + FFS: with smaller range(s) for differential L1-RSRP than legacy
  + FFS: step size(s) for absolute L1-RSRP, step size(s) for differential L1-RSRP, range(s) for differential L1-RSRP

|  |  |
| --- | --- |
| Company | Comments |
| FL | I don’t see motivation to have smaller quantization value.  There is no simulation to support it. |
| OPPO | Fine to have a study on Option 2. |
| MediaTek | We still support adding normalized L1-RSRP as one option. Let’s see below example, assuming legacy quantization and the model input is the L1-RSRP normalized in linear level per measurement instances (normalized by the largest measured power in the current measurement instances). The resulting possible model input values can be simplified as the following table:   |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | | Differential RSRP (dB) | 0~-2 | -2~-4 | -4~-6 | -6~-8 | -8~ -10 | -10~ -12 | -12~ -14 | -14~ -16 | … -28~  -30 | | Normalized input value | 1 | 0.63 | 0.4 | 0.25 | 0.16 | 0.1 | 0.063 | 0.04 | … 0.001 |   (\*normalized input value = 1/(10^(-0.1\*Diff\_RSRP\_dB)))  These are all the possible model input values. We can observe that the values are sparse at high range (>0.5) and dense at low range (<0.1). Note that this legacy method will use two-third of its quantization range for the input range between 0.001 and 0.1, and only one-third for the range between 0.1 to 1 (90% of its input range). On the other hand, quantizing the normalized results between 0 and 1 evenly using 4 bits can solve this issue and give more finer granularity to the high input range.   |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | | 1- x/16, x = | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | … 15 | | Normalized input value | 1 | 0.94 | 0.88 | 0.81 | 0.75 | 0.69 | 0.625 | 0.57 | … 0.063 |   We don’t know which input arrangement (which table) is more meaningful for AI/ML model. But in our view, the 2nd table is more reasonable since the AI/ML model should be sensitive to high input range, not low input range. Also, there are some supports on reducing the differential power range or data omission based on a threshold, meaning it’s a common understanding that higher power (higher input range) is more useful. Therefore, we prefer to keep the normalized option at this stage. |
| Hw/HiSi | Ok. |
| Intel | OK. |
| InterDigital | Fine for supporting option 1 but prefer to change the bullet for FFS as “FFS: other quantization of the reported RSRP”. |
| TCL | We agree with InterDigital. |
| CEWiT | Ok |
| CATT | We support further study option 2.  And there are some simulation results captured in TR 38.847, copied as following:   |  | | --- | | At least for BM-Case1 for inference of DL Tx beam with L1-RSRPs of all beams in Set B,  - Evaluation results from 4 sources show that, with 1dB quantization step for the absolute L1-RSRP of the best beam and 4dB quantization step differential L1-RSRP report with the existing quantization range, less than 5% beam prediction accuracy degradation in terms of Top-1 beam prediction accuracy compared to unquantized L1-RSRPs of beams in Set B.  - Same quantization scheme is used for the input data for training and inference. | |
| NEC | Ok, we see the motivation at least for BM-case2 if many historical results need to be reported. |
| QC | OK |

## 4 Configuration 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  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) |

|  |
| --- |
| 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 contributions:

|  |  |
| --- | --- |
| Company | Proposals |
| Futurewei [1] | ***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.*** |
| Ericsson [2] | Proposal 4 For UE-sided model at least for BM Case-1, CSI-ReportConfig is used for the configuration of inference results reporting, further study the following options for alternative 2  • Set A/B is indicated in the CSI Resource,  • Two CSI resource sets are used, one for set A and one for set B,  • Multiple resource sets are used for set A/B  Proposal 5 For UE-sided model at least for BM Case-1, regarding CSI-ReportConfig for the configuration of inference results reporting, down prioritize alternative 1 and 4  Proposal 6 For UE-sided model, conclude that the association/mapping of beams within Set A and beams within Set B provides UE with information of which beams that are part of each respective Set, and a new information element is required for such information.  Proposal 7 For UE-sided models, conclude that NW should determine which beams that are part of Set A and set B,  Proposal 8 For UE-sided models, regarding set A/B, prioritize a fixed set B/A over training and inference. Hence, down prioritize a variable set B.  Proposal 9 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 |
| HW/HiSi[3] | For UE-sided model training/monitoring, RAN1 to discuss the potential mechanism to enable the UE to perform CSI measurements on larger sizes of beam set(s) for Set A  • Set A consists of multiple resource sets each with legacy size (up to 64) of resources.  • Set A consists of one resource set with increased size of resources, e.g., 256.  Proposal 12: For UE-sided 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.  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. |
| Intel [4] | Proposal 3: For a UE-side AI/ML model, for BM-Case 1/2, set A should be 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. Set B may be configured via an association to set A or independently.  Proposal 4: 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. |
| Spreadtrum [7] | ***Proposal 9: For the configuration of inference results reporting for UE-side model, support Alt 1 or Alt 2.*** |
| Samsung [8] | **Proposal 7. For UE-sided model at least for BM Case-1, for the configuration of inference results reporting using *CSI-ReportConfig*, support the configurability between Alt 1 and Alt 3.**   * **Alt 1: one *CSI-ResourceConfigId* is configured for Set B**   + **the size of Set A is configured by the *CSI-ReportConfig*.** * **Alt 3: two *CSI-ResourceConfigId* s are configured for Set A and Set B separately** * **The resources configured in Set B can be periodic, semi-persistent or aperiodic.**   + **FFS: For BM-Case2, the configuration of aperiodic resources for Set B to facilitate measurements for multiple past time instance.**   **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.** |
| Vivo [9] | **Proposal 21: Due to significant performance degradation for mismatch pattern for model inference with UE-side model, UE can recommend preferred Set B patterns which were trained during the UE-side model training phase.**  **Proposal 22: For model inference with UE-side model, support to configure beam subset restriction of Set A associated with predicted beam report to address interference issue.** |
| InterDigital [10] | ***Observation 1:*** *Feasibility of Alt 1: one CSI-ResourceConfigId is configured for Set B is doubted as the gNB vendors are severely negative on providing the NW-sided additional conditions for constructing Set A.*  ***Observation 2:*** *Alt 2: one CSI-ResourceConfigId is configured for both Set A and Set B is beneficial as Alt 2 supports* *both scenarios (i.e., not transmitted Set A and transmitted Set A with larger periodicity) with minimum specification impacts.*  ***Observation 3:*** *Motivation of Alt 3: two CSI-ResourceConfigIds are configured for Set A and Set B separately is not clear as different channel/interference measurements, resource type or BWP ID are not needed.*  ***Observation 4:*** *Benefit of 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 is not clear as different CSI-ReportConfig needs to be supported considering different Set B even for an identical Set A.*  ***Proposal 1****: Support Alt 2: one CSI-ResourceConfigId is configured for both Set A and Set B.*  ***Proposal 2:*** *Support one of the following options for Set A and Set B configuration.*   * *Option 1: RS resources for Set A including Set B are configured in a RS resource set within a CSI-ResourceConfig.* * *Option 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 within a CSI-ResourceConfig.*   ***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.*   ***Observation 16:*** *For a UE-sided model, configuration information associated to a Set B can help the UE in ensuring consistency between training and inference.*  ***Observation 17:*** *Set B selection as a subset of the best measured beams for input to the gNB-sided AIML model achieves a higher beam prediction accuracy than reporting a uniformly or randomly selected subset of the measured beams.*  ***Proposal 14:*** *Indicating configuration information associated with Set B to UE should be supported.*  ***Proposal 15:*** *Support reporting of UE selected Set B based on a rule (e.g., subset of best measured beams).* |
| 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 RS configuration enhancement at least for the case that Set B and Set A are the same:**   * **Configure one RS set for measurement of Set B beams within the measurement window;** * **Configure multiple RS sets for measurement of Set B beams within the measurement window.**   **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.***   **Proposal 18：For UE side data collection, consider the following options for RS overhead reduction in network:**   * **Option 1: NW sends common RS configuration to different Ues for UE-side data collection;** * **Option 2: UE requests the preferred RS for data collection from NW pre-configured RS.**   **Proposal 19：For UE side data collection, the association of Set B and Set A can be defined based on following methods:**   * **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.*** |
| China Telecom [13] | ***Proposal 3: For UE-sided model, the UE can request the required/preferred measurement beams Set A/Set B via higher-layer signalling.***  ***Proposal 5: For the UE side model, during the beam inference phase, it supports the UE to request the corresponding Set A/Set B via higher-layer messages and reuse the TCI mechanism to activate/deactivate the corresponding set.*** |
| CMCC [14] | **Proposal 4: Regarding data collection for UE-side model of BM-Case1 and BM-Case2, the configuration method of set A for NW-side model can be reused.**  **Proposal 5: For UE-sided model at least for BM Case-1, *CSI-ReportConfig* is used for the configuration of data collection for training**   * **FFS on the details in the *CSI-ReportConfig*, at least considering:**   + **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***     - **indication of association of Set A and Set B can be based on RS ID or bitmap**   + **Option 2: two *CSI-ResourceConfigId* s are configured for Set A and Set B separately**     - **indication of association of Set A and Set B can be based on RS ID**   **Proposal 9: 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***     - **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**     - **indication of association of Set A and Set B can be based on RS ID** |
| Lenovo [16] | **Proposal 3: An indication is needed for the UE to select proper AI/ML model for AI/ML inference for a CSI-ReportConfig with AI/ML inference.**  **Proposal 4: Associate a prediction beam Set A for a measurement beam Set B for the UE to determine the Set A beams for a CSI-ReportConfig with AI/ML inference and the Set A beams are not needed to be explicitly configured for the CSI-ReportConfig.** |
| NVIDIA [17] | **Proposal 2: For BM-Case 1, at least introduce specification support for using L1-RSRP measurement based on Set B of beams as AI/ML model input.**  **Proposal 3: For BM-Case 2, introduce specification support for associating Set A of beams with Set B of beams.** |
| LGE [18] | **Proposal #3: Support reporting of UE assistance information for determining Set A, e.g., UE to report preferred Set A among candidate beams of Set A.**  **Proposal #7: Support Alt 1 or Alt 4 for Set A and Set B configuration for UE-sided AI/ML model.**  **Proposal #8: 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 #9: 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** |
| Panasonic [19] | **Proposal 4: UE-sided model inference, support two *CSI-ResourceConfigId* s are configured for Set A and Set B respectively.**  **Proposal 5: UE-sided model inference, support that a measurement window can be configured with the measurement resource set.**  **Proposal 7: Support mapping/association of beams within Set A and beams within Set B based on QCL relationship.** |
| Fujitsu [20] | ***Proposal 11:***   * *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)* |
| Xiaomi [21] | ***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 3-8: For model inference of UE-side model, in addition to the association ID to indicate the set A implicitly, a list of RS ID can be configured to indicate the mapping between the RS ID and the beam ID for NW side flexibility.*** |
| NEC [22] | Proposal 7: For triggering/initiating data collection at UE side for UE-side AI/ML model,  − Option 1: data collection initiated/triggered by configuration from NW is preferred, 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 16: 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 21: 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.*** |
| GOOGLE [23] | ***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.*** |
| ZTE [24] | ***Proposal 16: 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 17: 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 18: 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. => FL Where to provide such configuration?***  ***Proposal 21: 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.*** |
| ETRI [27] | Using this method, a single Set A defined CSI-ResourceConfig can also be associated with multiple CSI-ResourceConfigs defining different Set B beams.  **Proposal 5: Support the new ID for the association of Set A and Set B.** |
| Rakuten [28] | 1. ***RAN1 agrees the use case where, in a UE-sided model, beams of Set A cannot be always assumed to be transmitted before the prediction.*** 2. ***RAN1 agrees to define Set A to Set B association differently for transmitted and un-transmitted beams before prediction.*** 3. ***RAN1 discusses and agrees a solution to associate and identify un-transmitted beams predicted by UE in a UE-sided model.*** |
| OPPO [29] | ***Proposal 20: 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 21: For UE-side model inference of BM-Case2, configure Set A and/or Set B on a per time instance basis.***  ***Proposal 22: For UE-side model, two separate CSI-ResourceConfigIds can be configured within a CSI-ReportConfig (Alt 3).*** |
| Fraunhofer [30] | **Proposal 8: Set A of beams can be configured using the CSI reporting framework. FFS for further enhancements.**  **Proposal 9: Explore the possibility of using the resources configured for radio link monitoring for model monitoring purposes with respect to CSI beam reporting.**  **Proposal 10: Support a model monitoring configuration that allows for collecting data for model training and monitoring of inactive models.**  **Proposal 12: 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.** |
| Nokia [31] | **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.** |
| DoCoMo [32] | **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** |
| Sharp [33] | **Proposal 3:** For UE-sided model, for the details in the *CSI-ReportConfig*, further consider the followings:   * 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   **Proposal 4:** For UE-sided model, the association between Set A and Set B can be based on: In one *CSI-ReportConfig*, resources for Set A and resources for Set B are configured as separate resource sets with additional signalling to indicate the association.  **Proposal 5:** For triggering data collection at UE-side AI/ML model, support that UE sends a request for data collection.   * FFS: contents and container of the request. |
| MediaTek [34] | ***Proposal 13: For data collection, for report and resource configurations, consider the following options:***   * ***Option1: Set A and Set B resources are configured in one report configuration,*** * ***Option2: Set A and Set B resources are configured in multiple report configurations***   + ***how UE identify the corresponding report configurations for the current data collection instance.*** |
| KT[35] | ***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*** |
| Ruijie network [36] | **Proposal 7:** For UE-sided model at least for BM Case-1, *CSI-ReportConfig* is used for the configuration of inference results reporting. In the *CSI-ReportConfig,* support Alt 2:   * + Alt 2: one *CSI-ResourceConfigId* is configured for both Set A and Set B |
| ITL [38] | ***Proposal 5: 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 6: 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***  ***Proposal 14: 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*** |
| Qualcomm [37] | ***Proposal 7***  ***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)***  ***Proposal 12***  ***For beam prediction of UE-side AI/ML models, support mechanisms to identify an AI/ML functionality by associating it to a CSI report setting (including reference signal configurations), wherein activation/deactivation of an AI/ML functionality is indicated by activation/deactivation of the associated CSI report setting option.*** |

#### Issue #1: Configuration for inference results reporting

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
      * Opt 1 implicitly obtained via association ID
      * Opt 2: link to a *CSI-reportConfig* for Set A with “None” of the report
  + *Supported by (7): Spreadtrum, Samsung, LGE, Fujitsu?, ZTE, Nokia, KT*
  + *Deprioritize by: Ericsson*
  + 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(11): Futurewei, Ericsson, Spreadtrum, Interdigital, CMCC (RS ID or bitmap) Fujitsu, Fraunhofer, Nokia, Sharp (association with additional signaling), KT, ruijie*
  + Alt 3: two *CSI-ResourceConfigId* s are configured for Set A and Set B separately
  + *Supported by(10): Futurewei, Huawei, Samsung, CMCC, Panasonic, ZTE, OPPO, DoCoMo, Sharp, ITL*
  + 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 LGE* 
      * *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*
      * *Fujitsu?*
      * *Nokia (for data collection/performance monitoring)*
      * *DoCoMo*
    - *Deprioritize by: Ericsson,*
  + 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.

##### **(High)Questions to answer**

A: Do you think RS resources for full set of Set A will be configured to UE anyway if the AI model is activated? If yes, when?

* e.g., for monitoring?
* e.g., for training data?
* e.g., for inference? *FL: if UE will be configured for Set A for other purpose, e.g., training, whether the RS resources for full set A is needed in Report configuration for inference?*

B: Do you think common design is needed for the case Set A is different from Set B, and Set A is a subset of Set B?

|  |  |
| --- | --- |
| Company | Proposals |
| Hw/HiSi | Question A:  Full Set A will be configured for measurement and reporting for training and for monitoring. This is needed to obtain the label. For inference, full Set A will be configured for reporting, so that the UE can select the best beams from the inference output. It seems the knowledge of the full Set A is needed, but the UE behaviour maybe different depending on the purpose.  Question B:  A common design may not be needed, but helpful. For Set A, we need to configure the resources and the behaviour (measurement and/or reporting).  When Set A and Set B are different, Set B will be measured. The UE should know the associated between beam in Set B and Set A. And the Set A will be reported. Mapping of Set B to Set A is not needed.  When Set B is a subset of Set A, set B will be measured. Set A can be configured separately. Set A can be linked with a corresponding Set A, the beams of Set B can be indicated with a bitmap in the configuration of Set A and Set A is configured for reporting. |
| TCL | A: Full set of Set A should be configured to UE when training and inference, but not for monitoring. Monitoring on a sub set of Set A will still reflect the performance.  B: It is not necessarily to limit Set A as a subset of Set B. For example, we can predict CSI-RS beams using SSB beams. |
| Vivo | A: Yes. Set A should be configured among training, monitoring and inference procedure.  B: the typo for the last sentence, it needs to be changed as “Set B is a subset of Set A”. |
| MediaTek | Question A: We would like to know the purpose of this question. Does this mean NW may not need to configure Set A information for inference reporting if RS resources of full Set A is any way configured to UE during inference/monitoring/training? RS resources of full Set A may not be configured to UE during inference and monitoring. NW may configure RS resources of full Set A to UE during training. However, even at the same cell, the RS resource ID for each Set A beam can be different for different training data collection instances.  Question B: We prefer a common design |
| NTT DOCOMO | A: Yes, for monitoring and training purposes.  B: We do not find the strong motivation to differentiate “Set A is different from Set B” and “Set A is a subset of Set B”. In that case, the common design should be made. |
| QC | A: Yes, configuring Set A resources for all the purposes (including inference) would help with leveraging and building on top of the existing CSI framework, with minimal changes.  B: if “common design” is meant to refer to the above alternatives, yes. Also please add QC to supporters of Alt. 2. |
| CATT | A: Yes for training, and No for inference and monitoring |
| ETRI | A: Full Set A will be configured for training data. But, for performance monitoring, only subset of set A can be configured.  B: We think that common design may be used. |
| ZTE | A: For monitoring or training, the configuration of Set A is needed since the UE would measure resources of Set A to get the model label data. For inference only, the configuration of Set A is not needed since the UE would not measure resources of Set A. However, if a UE is responsible for both inference and other purposes, e.g., training, anyway Set A needs to be configured to the UE. And in that case, the inference report only needs to link to the configured Set A for the provision of information about Set A, and thus the additional configuration overhead is minimal.  B: Not necessary. There are some significant differences between the two cases. For example, if Set A is different from Set B, both resources of Set A and Set B need to be explicitly configured to the UE. However, if Set B is a subset of Set A, only resources of Set A need to be explicitly configured to the UE, and resources of Set B can be indicated as a subset of Set A to save configuration overhead. |
| Panasonic | Question A: Yes. At least for model training.  Question B: Yes, a common design is needed. |
| Xiaomi | A: Yes for training purposes. For monitoring, if the metric is L1-RSRP difference, a subset of set A is sufficient. For inference, the associated ID can be used to indicate the information of set A implicitly. But in order for NW flexibility, the RS ID of set A can be configured in addition so the mapping between beam ID and RS ID can be changed at NW.  B: for inference, we think common design is ok. But for training, signalling overhead can be reduced if set B is a subset of set A by different design. |
| Intel | **A:** No, not for inference. Resources associated with set A for a UE sided model may not need to be physically transmitted as reference signals for the UE to make measurements on, except for the case of training data collection or some instances of model performance monitoring.  **B:** While common design may always be preferable, we do not need to consider that as a target already before the details of the configuration of sets A and B are decided.  On the options, we also support Alt 4;   * + 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 LGE, Intel*    + *One or more separate resource set(s) for Set A can be configured outside of CSI-ResourceConfig* |
| NEC | **Question A:** Not always. For training it may be needed. For monitoring only a subset is needed, for inference, it may not be needed.  **Question B:** Not necessary, if set B is a subset of set A, maybe only set A is configured and a bitmap is for indicating the set B. if set B and set A are different, two configurations are needed |
| New H3C | **A: for training at least** |
| InterDigital | **A: Yes**  **B: Yes** |
| Ericsson | **Question A:** Yes for inference, they can be configured but not transmitted by the NW, only the report of such resources can be performed.  Yes for monitoring and training.  **Question B:** Need some clarification on what a common design is. |
| SPRD | A: Full set of Set A should be configured to UE at least for training.  B: It is not necessary. |
| LG | A: If full set of Set A is configured, it can be for training and monitoring. For inference, resources in Set A may or may not be configured for inference result report. |
| Fujitsu | Question A: Yes.  Question B: Common design is preferred. |
| Google | Question A: No. The overhead is too much. In addition, what’s the benefit for this feature if this is mandated?  Question B: No. |
| CMCC | A: for the full set configuration, at least for training and inference, the full set of Set A should be configured. Different size of set A may induce different usage of the AI model, which may lead mismatching between training and inference.  B: a common design may be helpful for some cases. But this is only for the indication of set A and Set B, it is not necessary to limit to the same design when the set B is a subset of set A and the set B is not the subset of Set A. |
| CAICT | A: Yes. Set A should be configured to UE.  B: Yes. |
| Lenovo | A: It’s at least needed for model training. Whether it’s needed for performance monitoring may depend on the metric for monitoring.  B No. |
| OPPO | A: for training and inference, full Set A should be configured to UE.  B: Strive for unified design for both cases. |
| Apple | B: depending on consistency discussion, it may be beneficial to have different treatments for those two cases. In any cases, we can discuss the intended UE behaviors and NW operation first, then come to the detailed design. |

#### Issue #2: Request for RS resource for Set A

*FL: Several companies propose to support UE request RS resources of Set A*

##### **(low)Proposal 4.2A:**

Support UE to request for RS resources of Set A, FFS on details

|  |  |
| --- | --- |
| Company | Proposals |
| TCL | We do not support this proposal as it disorders the resource configuration procedure. |
| Vivo | Support, but do we need to illustrate the intention of the request? |
| MediaTek | We think this can be addressed by NW and UE align the “associated ID” (if agreed) for the AI/ML model |
| NTT DOCOMO | Support. |
| CATT | Ok to discussion. |
| NEC | OK to discuss. |
| Ericsson | We think the need should be further studied, hence.  Updated Proposal 4.2A. Further study the need for UE to request for RS resources of Set A |
| LG | Support. |
| Fujitsu | What’s the use case for this proposal? This should be clarified. |
| Google | Support |
| CMCC | Same comments as vivo that further illustration or details of requesting of set A RS resources needs clarification. From our understanding, the transmission of RS resources can be requested for the measurement at least for the TCI states, when the UE have never measured the reference RS of the TCI state and which belongs to the Set A |
| Lenovo | Support |
|  |  |

#### Issue #3: Configuration for the measurements of past time instances for BM-Case 2

##### **(low)Question to answer:**

How to configure for the measurement of past time instances for BM-Case 2

|  |  |
| --- | --- |
| Company | Proposals |
| FL | Any view? |
| TCL | Opt 1: Measurement/observation (time) window, UE by default measures all the RS configured within the window.  Opt 2: Measurement/observation (time) window + number of measurements.  Opt 3 Measurement/observation (time) window + time interval of measurements.  Opt 4: Measurement/observation (time) window + pattern of measurements.  Opt 5: Number of measurements + time interval of measurements. |
| Vivo | Define a new reporting type, and then the UE can know which is configured for the measurement of past time instances for BM-Case 2. Furthermore, the observation window (similar to Rel-18 prediction) can be considered. |
| NTT DOCOMO | We can simply reuse R18 type II ignali CSI reporting for measurements of past time instances, unless some enhancements are identified necessary. |
| CATT | Option 1: configure one RS set with periodicity X of measurement window and the periodicity Y of RS transmission and the number of RS within a measurement window.  Option 2: configure multiple RS sets for measurement of Set B beams within the measurement window. |
| NEC | We think the configuration of measurement window or the number of measurement instances can work. |
| Fujitsu | Rel-18 MIMO design could be starting point. |
| Google | We can use similar approach as R18 CSI |
| CMCC | Opt 1 Measurement/observation (time) window length+ time interval of measurements.  Opt 2:first time occasion + number of measurements + time interval of measurements. |
| Lenovo | Rel-18 CSI report configuration framework is the starting point. |
|  |  |

## Inference result report for UE-sided model report

### Issue #1: Content of inference results for UE sided model

* 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
  + Yes:
    - Ericsson [2] UE report of an uncertainty of a predicted beam can be used by the NW when configuring a subsequent Top-K measurements. It can also enable the NW to fallback to a legacy beam management method in case the uncertainty is high
    - Huawei/HiSi [3] For the content in the report of the AI/ML model inference at the UE-side,
      * For the probability information of the beam IDs, consider following solutions:
        + Opt 3-1: Reporting the probability information of predicted Top-K beams.
        + Opt 3-2: Reporting the selected beams determined by UE based on probability threshold. *=>FL: I think this is to be used to select opt 1, top K beam*
    - ZTE [24]
    - GOOGLE [23]: The probability information and the confidence information are useful for the network to determine whether to perform TCI switching based on the beam prediction result or not.
    - OPPO [29] For BM-Case1 and BM-Case2 with UE-side model, the contents of inference results could include (Opt 3) probability information of predicted Top K beam(s).
    - Samsung [8] the probability information could also be useful since the probability can reflect beam prediction accuracy in some extend
    - LGE [18] 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.
    - NEC [22] 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.
    - Sony [15] considering the content in the report of inference results, we support Options 1, 2, and 3.
      * Supporting the sum probabilities of predicted K beams exceeding a predefined threshold for being the Top beams can serve as the criterion.
    - Nokia [31] The probability information of predicted Top K beam(s) can be defined, for example, as the probability of each beam in Set A to be the Top-1 or Top-K beam(s) as described in clause 6.3.1 of TR 38.843. This information can be used by the gNB to assess the quality and reliability of the prediction reported by the UE and may have several uses, e.g. TCI activation among others. **Probability information shall be the probability of the beam to be the Top 1.**
    - Fraunhofer HHI, Fraunhofer IIS [30] For UE-sided models, for inference, examine whether and how to report confidence of prediction.
    - KDDI [32] Probability information and confidence information should be defined. **Information based on output probabilities in AI/ML model classification**
    - Indian Institute of Tech (M), IIT Kanpur [37] When the AI/ML model output is the beam ID, the beam report should contain (1) beam IDs of the predicted Top *K* beam(s) among a set of beams and (2) probability information of the predicted Top *K* beam(s) (Opt 3 [1])
  + No:
    - Futurewei [2] it is difficult to define and test these new metrics
    - CATT [10]: but whether the probability is reliable is unknown
    - ITL [38] We do not believe that additional information beyond the specified ones (e.g., Top-K predicted beam ID/L1-RSRP), such as probability information or confidence information, needs to be reported to the base station.
  + Others:
    - DoCoMo [32] 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.
* 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
  + Yes:
    - Ericsson [2] UE report of an uncertainty of a predicted beam can be used by the NW when configuring a subsequent Top-K measurements. It can also enable the NW to fallback to a legacy beam management method in case the uncertainty is high
    - GOOGLE [23]: The probability information and the confidence information are useful for the network to determine whether to perform TCI switching based on the beam prediction result or not.
    - NEC [22] Support UE to report confidence information associated with the predicted L1-RSRPs, 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%).
    - Fraunhofer HHI, Fraunhofer IIS [29] For UE-sided models, for inference, examine whether and how to report confidence of prediction.
    - Qualcomm [37] 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.
  + No:
    - Huawei/HiSi [3]: The necessity of confidence information of the RSRP (Opt 4) of predicted Top-K beams is not clear.
    - Futurewei [2] it is difficult to define and test these new metrics
    - CATT [10] both the definition and the scheme to obtain the confidence/probability information related to the predicted result are still FFS
    - Nokia [25] Do not support Opt.4.
    - ITL [38] We do not believe that additional information beyond the specified ones (e.g., Top-K predicted beam ID/L1-RSRP), such as probability information or confidence information, needs to be reported to the base station.

### Issue #2: Definition of RSRP ofpredicted Top K beam(s)

|  |
| --- |
| 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 |

|  |  |
| --- | --- |
| Company | Proposals |
| Futurewei [1] | ***For Rel-19 AI/ML-based BM, 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, when applicable, support Option B***   * ***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.*** |
| Intel [4] | Proposal 15: 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. |
| Samsung [8] | **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.** |
| Vivo [9] | Proposal 26: For model inference with 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.  Proposal 27: For model inference with UE-side model, confirm the working assumption that for report content of inference results for UE-sided model for BM-Case 2, the RSRP of predicted beam(s) in the report of inference results, is the predicted RSRP, where the predicted RSRP is based on AI/ML output. |
| Interdigital [11] | For report content of inference results for UE-sided model for BM-Case 1, the difference between ‘Option A: Predicted RSRP’ and ‘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’ is not clear. Specifically, Option A should be further clarified including details on when/how predicted RSRP needs to be reported.  ***Proposal 12****: Option A “Predicted RSRP” should be further clarified.* |
| CATT [12] | **Proposal 10: 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.** |
| CMCC [14] | **Proposal 10: 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.** |
| Lenovo [16] | **Proposal 8: Support to report the predicted RSRP for the reported beams in a beam report with AI inference for BM-Case 1 .**  **Proposal 9: Confirm the following working assumption**  **Working Assumption**  **For report content of inference results for UE-sided model for BM-Case 2, the RSRP of predicted beam(s) in the report of inference results, is the predicted RSRP, where the predicted RSRP is based on AI/ML output.**  **Proposal 14: 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.** |
| LGE [18] | **Proposal #15: 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** |
| Panasonic [19] | **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.** |
| Fujitsu [20] | ***Proposal 13:***   * *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.* |
| Xiaomi [21] | ***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.*** |
| NEC [22] | ***Proposal 20: 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.*** |
| GOOGLE [23] | * ***The reported RSRP should be defined as the predicted RSRP based on a reference transmission power*** |
| ZTE [24] | 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. |
| Hyundai Motor Company [26] | * **Discuss whether not only predicted RSRP but also measured L1-RSRP are considered to select Top K beams.** * **Discuss whether/how UE reports difference between measured L1-RSRP and predicted RSRP if the beam is configured for corresponding measurement.** |
| Fraunhofer [30] | **Proposal 3: For UE-sided models, for inference, support reporting of predicted RSRP values.**  **Proposal 20: Use of a predicted beam that is not measured/received by the UE for beam indication is supported.** |
| Nokia [31] | * + **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.** * **FFS: whether measured beam related quantities (CRIs, L1-RSRP) of Set B can be configured to report within the same beam report.** |
| DoCoMo [32] | **Proposal 9: Predicted RSRP can be defined as the expected RSRP that UE would measure based on the corresponding RS at the associated time instance.**  **Observation 2: From RAN1 specification perspective, reported RSRP does not need to be differentiated into predicted RSRP and measured RSRP.** |
| Ruijie Network [36] | **Proposal 6: Confirm the working assumption:**  ~~Working Assumption~~ 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 of the status:

* Option A: Predicted RSRP
  + *Supported by: Intel? Samsung, CATT, CMCC, Lenovo, LGE, google Fraunhofer*
* 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
  + *Supported by: Ericsson, Intel?, Samsung, CATT, CMCC, xiaomi, NEC, ZTE*
  + *Intel, Hyundai?: Support differentiation.*
  + *CATT, CMCC: configurable*
  + *Lenovo: indicated which*
  + *LGE: up to UE implementation*
  + *Fujitsu: based on performance? [Unclear]*

### Issue #3: Inference report for BM-Case 2 for UE sided model

|  |  |
| --- | --- |
| Company | Proposals |
| Ericsson [2] | Proposal 11 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. |
| Huawei/HiSi [3] | 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.  For BM-Case 2 with a UE-side model, consider to introduce one DCI indicating multiple resource sets subject to multiple time instances as the observation window under A-CSI. |
| Spreadtrum[7] | *Proposal 7: Reporting multiple past time instances in one reporting instance for BM-Case2 is not needed.*  *Proposal 8: For BM-Case2, implicit report of time information should be supported.* |
| Samsung [8] | 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) |
| Vivo [9] | Proposal 25: For model inference with UE-side model, support time stamp information in beam content for BM-Case2. |
| Apple [10] | **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.** |
| InterDigital [11] | ***Observation 13:*** *For BM-Case 2 with a UE side AI/ML model, reporting information of multiple future time instances in one report reduces latency as gNB does not have to wait for UE reports in queue.*  ***Proposal 11:*** *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.* |
| CMCC [14] | **Proposal 11: Confirm the following working assumption:**  **For report content of inference results for UE-sided model for BM-Case 2, the RSRP of predicted beam(s) in the report of inference results, is the predicted RSRP, where the predicted RSRP is based on AI/ML output.** |
| Sony [15] | **Proposal 4 : For BM-Case2, especially for UE-side model, gNB and UE should align the timestamps of the future N time instances.**  **Proposal 5 : For use case 2, gNB can indicate to the UE the duration of the time window for collecting input data based on the characteristics of time channel.** |
| Lenovo [16] | **Proposal 10: Study the mechanism for the UE to determine the measurement window and a prediction window for BM-Case2.** |
| LGE [18] | **Proposal #12: 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 #13: For temporal DL Tx beam prediction, information on time-variation of RSRP can also be included in the report.**  **Proposal #14: 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** |
| Fujitsu [20] | * *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.* |
| Xiaomi [21] | ***Proposal 4-2: Support to report the measured beam information of the history measurement time instance for UE side model inference in BM Case 2 if set B equals to set A.***  ***Proposal 4-3: 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 4-4: 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.*** |
| OPPO [29] | ***Proposal 13: For BM-Case2 with UE-side model, the timestamp of future time instance(s) could be implicitly reported to NW.***  ***Proposal 14: For BM-Case2 with UE-side model, support to indicate multiple beam indications for future time instances with one-shot beam indication.*** |
| Nokia [31] | **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.** |
| DoCoMo [32] | **Proposal 10: Reuse multiple CSI-RS resource occasion association for Rel-18 type II ignali CSI reporting, if applicable.**  **Proposal 12: Support the following payload overhead reduction for UE reporting of predicted results at multiple time instances.**  **・Differential RSRP representation from reference RSRP at different time instance** |
| Sharp [33] | **Proposal 7:** For UE-sided model, for BM-Case 2, for content in the report of inference results, support providing time interval information between multiple future time instances.  **Proposal 8:** For beam reporting information of BM Case 2 model, the range of past time instances or a time window should be defined.  **Proposal 9**: For timing window in BM Case 2 model, take the time stamp configuration as the starting point to indicate the timing of inferred results in one report. |
| ITL [38] | ***Proposal 16: 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*** |
| Vivo [9] | Proposal 30: For model inference with UE-side model, support time domain compression of beam resource indication to further reduce report overhead with a report including results of multiple occasions.  Proposal 31: For model inference with 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. |
| GOOGLE [23] | ***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*** |
| ETRI [27] | **Proposal 8: For UE-sided model, for temporal domain beam prediction, support the CSI report format with the optimal K beam information along with RSRP information from multiple time instances.** |

### Issue #4: FFS on beam information

|  |  |
| --- | --- |
|  |  |
| Ericsson [2] | Proposal 12 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) |
| Huawei/HiSi[3] | *Proposal 18: 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.* |
| H3C[5] | **Proposal 6: For the UE side model, in the beam inference phase, it is suggested that the UE use L1 signaling to report the top-K beams based on CRI and RSRP. The maximum number of beams reported should be consistent with the number of measurements in Non-AI/ML Beam Management.** |
| Samsung [8] | **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.** |
| Vivo[9] | **Proposal 2: For data collection procedure with UE-side model, use RS (SSB or CSI-RS) resource ID to configure Set A and/or Set B.** |
| Fujistu[20] | * + *The beam information could include CRI/SSBRI and CC ID.* |
| GOOGLE [23] | * ***Beam information is defined as a beam indicator (BI) indicating one of the beams from a configured codebook*** |
| ZTE [24] | * ***The beam information is CRI or SSBRI for BM-Case1, FFS for BM-Case2*** |
| OPPO [29] | ***Proposal 11: For BM-Case1 and BM-Case2 with UE-side model, clarify the beam information on predicted Top-K beam(s) as SSBRI/CRI associated with Set A.*** |
| Nokia [31] | * + **Beam information refer to CRIs corresponding to Set A** |
| Qualcomm [37] | **Proposal 5**  **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.** |
|  | **Proposal 7: Beam information on predicted top K beam(s) should be represented by CRI/SSBRI to follow the existing specification.** |

Beam information for predicted beam:

Alt 1: the CSI-RS resource indicator (CRI) and SSB resource indicator (SSBRI)

* Ericsson [3], Samsung [8], vivo [9], ZTE [7],
* OPPO [9], Nokia [25] FFS for predicted beam, SSBRI/CRI associated with Set A
* Fujitsu [19] The beam information could include CRI/SSBRI and CC ID.
* DoCoMo [35] Beam information on predicted top K beam(s) should be represented by CRI/SSBRI to follow the existing specification.

Alt 2: Beam information is defined as a beam indicator (BI) indicating one of the beams of Set A

* Samsung?? [8] (for predicted beam, defined a DL beam ID)

Alt 3: Beam information is defined as a beam indicator (BI) from a configured codebook

* GOOGLE [23]

### Other proposals

|  |  |
| --- | --- |
|  |  |
| Intel [4] | Proposal 17: 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. |
| *Samsung [8]* | **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.**  *FL: for option 1?* |

|  |  |
| --- | --- |
| Samsung [8] | **Proposal 15. For UE-side AI/ML model inference, support differential L1-RSRP reporting for predicted beams.** |
| Vivo [9] | Proposal 28: For model inference with UE-side model, support following report enhancement for overhead reduction:   Support enhancements on quantization range and quantization step   Support adaptive number of beams in a beam report |
| Lenovo[16] | Proposal 7: Support periodic beam report, semi-persistent beam report on PUCCH, semi-persistent beam report on PUSCH, aperiodic beam report with AI prediction. |
| NEC [22] | Proposal 18: Support selecting Top-K beam(s) according to some pre-defined rules (e.g., a sum probability of being Top 1 or Top K beam higher than a threshold, predicted L1-RSRP higher than a threshold) as the reported predicted beams. |
| GOOGLE[23] | * ***The selection of the “top-K” beams are up to UE implementation***   + ***Spec only defines the number of reported beams*** |
| ETRI [27] | **Proposal 6: Support the new CSI report format or indication field for model inference when an AI/ML model is located on the UE-side.**  **Proposal 7: 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.** |
| OPPO[29] | **Proposal 12: For BM-Case2 with UE-side model, study overhead reduction (e.g. enhanced differential L1-RSRP reporting) to facilitate multiple future time instances within one report.** |
| MediaTek [34] | ***Proposal 8: For UE-side model inference, study both periodic and aperiodic report and Set B resource configurations.***  ***Proposal 9: For UE-side model inference, same design of report and resource configurations can be used for BM Case1 and BM Case2***  ***Proposal 11: 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.*** |
| ITL [38] | ***Proposal 12: 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 13: For UE side model inference, existing specifications should be the baseline for the configuration or triggering CSI-RS/SSB of Set B***  ***Proposal 15: 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 15: 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\*** |

### 1st Round discussion

#### Issue #1: FFS on beam information

TBD, depends on configuration of Set A and Set B, and after for NW sided model

#### Issue #2: FFS on the definition of predicted Top K beam(s)

*FL’s comment: two issues need to be discussed*

*#1: how to define predicted Top K beam(s)*

*#2: how to convey ranking information*

##### **(low)Potential Proposal 5.2A:**

The predicted Top K beam(s) in the report for inference results for UE-sided model is defined as Opt A or Opt B (according to the AI model output):

* Opt A: beam(s) with Top K largest value(s) of probability of the beam to be the Top 1
* Opt B: beam(s) with Top K largest value(s) of predicted RSRP

FFS on how to convey the ranking information of Top K beam(s) at least for Opt 1: Beam information on predicted Top K beam(s) among a set of beams

|  |  |
| --- | --- |
| Company | Comments |
| FL | Predicted Top K beams needs to be defined, and somehow, specify in the spec.  For Opt 1, I think implicitly way, order = ranking  For Opt 2, since RSRP is reported, gNB is able to identify the ranking |
| HW/HiSi | For discussion it seems there is a relationship with the agreement that we made last meeting:  *Agreement*  *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, 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*   For Opt A from the proposal for example, we think there could be a relationship/conflict with the reporting by Option B from the agreement. For example, i) if a predicted Top-K has a measured L1-RSRP that would change its ranking inside the Top-K, what would be the reporting behavior? Or ii) if a predicted beam that is not top-K has a measured L1-RSRP that is larger than the predicted Top-K shall it be reported instead of a predicted Top-K? |
| TCL | Agree with the FL comments. |
| Vivo | We think main bullet is enough, the benefit or meaning of sub-option is unclear, for example, the Top K beam can be decided directly by AI model output) |
| ZTE | Support |
| NEC | We think it is up to UE implementation to choose K to report, there is no need and it is also very difficult to mandate K largest RSRP value or K largest probability. |
| Ericsson | In general ok with the direction, but how could RAN4 test such UE behavior? |
| SPRD | Agree with vivo. It can be decided by AI model output |
| Fujitsu | Does the proposal mean both options will be supported? |
| Google | We think the top-K can be up to UE implementation |
| CMCC | Ok. |
| Lenovo | Fine |

#### Issue #3: FFS on definition of reported RSRP

##### **(low)Proposal 5.3A:**

Confirm the following working assumption:

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

|  |  |
| --- | --- |
| Company | Comments |
| FL | No complains received. |
| TCL | Agree with the WA. |
| Vivo | support |
| NTT DOCOMO | From NW perspective, it does not matter whether predicted RSRP is based on output of AI/ML model or non-AI algorithm. Hence, we prefer to change the definition of predicted RSRP as the expected RSRP that UE would measure based on the corresponding RS at the associated time instance. If this definition is adopted, predicted RSRP also can include RSRP predicted by non-AI algorithm, which is preferable in terms of future compatibility. |
| CATT | Support the working assumption. |
| ZTE | Support |
| SPRD | Support. |
| LG | Support. |
| Fujitsu | Ok. |
| Google | Support |
| CMCC | Support the working assumption. Since in BM-Case 2, the RSRP is predicted for the future time instance, unless other cases would be identified. |
| Lenovo | Fine |

#### Issue #4: Inference report for BM-Case 2

##### **(low)Potential Proposal 5.4A:**

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)

|  |  |
| --- | --- |
| Company | Proposals |
| FL | This is a valid question to be discussed. Let’s collect views in 1st round.  Any other |
| TCL | The value of P depends on valid period of the predicted beams. For example, if N=2, each beam valid for 1 slot, then P=2; if the first beam valid for 1 slot while the second one valid for 2, then P=3.  We support Option 1 and 2. We think Option 3 is not that good, since model inference takes time and is diverse among Ues, this may lead to different configurations for different Ues. |
| Vivo | For the first sub-bullet, more clarification is needed for what means ” Each of the N future time instance(s) consists of P (P≥1) consecutive slots” |
| NTT DOCOMO | We think prediction future time instance determination can follow R18 type II ignali CSI where prediction window is determined based on the offset from reporting slot, unless some enhancements are necessary. |
| ZTE | The first bullet contradicts with the last second FFS regarding whether the separation between two adjacent time instance(s) are the same or not. Thus, we suggest the following revision.   * ~~Each of the N future~~ The time interval between two adjacent time instance(s) consists of ~~P (P≥1)~~ not lower than one consecutive slots   + FFS: How to determine ~~P~~ the number of consecutive slots   + FFS: If N>1, whether the time domain separation between two adjacent time instance(s) from the N future time instances are the same |
| NEC | We think Option 1 could work. |
| LG | I’m not sure why N future time instances include consecutive slots. The time instances can be reported via time stamp in the inference reporting. Then, there is no need to discuss time reference. |
| Fujitsu | Generally fine with the proposal. |
| Google | Support in principle |
| CMCC | For the 1st bullet, it should be clarify why do we need the P consecutive slots. For the determination of N future time instance, with the determination of the number N and time span between each time instance, we can have a clear definition of the N time instance. It is not clear to us why we need the P consecutive slots or the transmission occasions. |
| Lenovo | Generally fine. |

#### Issue #5: Support of Opt 3 and/or Opt 4 for inference report

Discuss together with performance monitoring.

#### Issue #6: Omission and quantization of the report

Discuss NW sided model first and then check whether can be apply for inference report for UE sided model.

### 5.2 2nd Round discussion

#### Issue #3: FFS on definition of reported RSRP

##### **(High)Proposal 5.3A:**

Confirm the following working assumption:

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

|  |  |
| --- | --- |
| Company | Comments |
| FL | No complains received. |
| TCL | Agree with the WA. |
| Vivo | support |
| NTT DOCOMO | From NW perspective, it does not matter whether predicted RSRP is based on output of AI/ML model or non-AI algorithm. Hence, we prefer to change the definition of predicted RSRP as the expected RSRP that UE would measure based on the corresponding RS at the associated time instance. If this definition is adopted, predicted RSRP also can include RSRP predicted by non-AI algorithm, which is preferable in terms of future compatibility. |
| CATT | Support the working assumption. |
| ZTE | Support |
| SPRD | Support. |
| LG | Support. |
| Fujitsu | Ok. |
| Google | Support |
| CMCC | Support the working assumption. Since in BM-Case 2, the RSRP is predicted for the future time instance, unless other cases would be identified. |
| Lenovo | Fine |
| OPPO | Okay. |
| Intel | OK |
| InterDigital | Fine |
| CEWiT | Support |
| NEC | Ok. |

## Beam indication

|  |  |
| --- | --- |
| Company | Proposals |
| Ericsson [2] | 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 |
| Huawei/HiSi[3] | Observation 6: For BM-Case 1 with a NW-side model, the legacy TCI framework and mechanism are sufficient for handling beam indication.  Proposal 16: 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 DCI.   + gNB updates/overrides the TCI state that is previously predicted before the corresponding future time instance. * Substantial impact on implementation complexity and RAN4 impact (e.g. an increased number of active TCI states). |
| Intel [4] | Proposal 17: 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. |
| Spreadtrum [7] | ***Proposal 5: 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.*** |
| Samsung[8] | **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.** |
| Interdigital [10] | ***Observation 5:*** *Update of existing beam indication framework is needed to enable indication of beams as TCI states for unmeasured beams as corresponding TCI states cannot be configured with physical RS IDs as QCL Type-D reference RSs.*  ***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.*   ***Observation 6:*** *When PDCCH/PDSCH are to be transmitted via a predicted best beam which is not measured, the UE cannot acquire QCL-related parameters, therefore a procedure to obtain QCL-related parameters for the unmeasured beam is needed.*  ***Observation 7:*** *Estimation of QCL-related parameters via neighbouring beams achieves Doppler shift estimation with error below 10 Hz for 90% of the time and RMS delay spread below 15 ns for 90% of the time.*  ***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.* |
| 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.** |
| Fujitsu [20] | * *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.* |
| Xiaomi [21] | ***Proposal 4-4: 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 [22] | ***Proposal 34: 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 35: For BM-Case 2, support to use one MAC CE or DCI to activate/indicate multiple (future) TCI states, and corresponding time period*** |
| GOOGLE [23] | ***Proposal 6: Support dynamic activation/deactivation of periodic TRS with regard to TCI activation/indication based on the predicted beam.***  ***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.*** |
| ZTE [24] | ***Proposal 22: For BM-Case2 (both UE-sided and NW-sided model), support to extend the Rel-17 TCI state activation/indication ignaling methods to activate/indicate N TCI states which are corresponding to N future time instances.*** |
| ETRI [27] | **Proposal 13: Support the method for representing the relationship between multiple beams and the predicted beam from UE-sided model.** |
| OPPO [29] | **Proposal 5: For BM-Case2 with NW-side model, enhance unified TCI framework to facilitate beam indication for multiple future time instances.** |
| **Fraunhofer [30]** | **Proposal 21: Study whether beam indication for multiple future time instances is required.** |
| **Nokia [31]** | **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 ignaling 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 17: For BM-Case2 with the NW-sided model, consider extending the Rel-17 TCI state activation/indication ignaling methods to activate/indicate N TCI states corresponding to future time N instances.** |
| **Sharp [33]** | **Proposal 10:** For NW-sided model and UE-sided model, enhance unified TCI state framework to support beam indication of multiple future time instances. |
| **MediaTek [34]** | ***Proposal 12: For AI/ML-based BM, at this stage, there is no further enhancement needed for beam indication based on unified TCI state framework.*** |
| **ITL [38]** | ***Proposal 10: 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 11: 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*** |

##### **Summary of the view for beam indication for multiple further time instances**

For BM-Case2 (both UE-sided and NW-sided model), study on whether/how to extend the Rel-17 TCI state activation/indication ignaling 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.

Yes(14): Spreadtrum, Samsung, Interdigital, CATT, Fujitsu?, Xiaomi, NEC, ZTE, ETRI, OPPO, Fraunhofer, Nokia, Sharp, ITL

No(2): Huawei/HiSi, MediaTek

### 6.1 1st Round discussion

#### Issue #1: Whether and how to support beam indication for multiple further time instances

##### **(low)Potential proposal 6.1A:**

For BM-Case2 (both UE-sided and NW-sided model), study on whether/how to extend the Rel-17 TCI state activation/indication ignaling 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.

|  |  |
| --- | --- |
| Company | Comments |
| FL | Let’s not stuck on whether, but study how together with whether.  I don’t see a big problem to study this with 14 vs 2 support |
| HW/HiSi | A discussion of this proposal, even it includes a “whether” would be time-consuming, that we may be able to spend on more critical aspects.This is not a critical proposal since the legacy framework works well. Could proponents or the FL please clarify for which use case there would be a potential benefit? As we see it, the only potential benefit would be to save overhead when Top-1 beams are predicted and when no PDSCH is scheduled. This is a corner-case, but inflicting significant implementation complexity and spec effort:   * Firstly, the overhead saving would only occur if no PDSCH would be scheduled, or there is infrequent DL traffic, since otherwise a DCI is anyway transmitted in which the TCI state can be indicated. * Secondly, 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. * Thirdly, 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. * Fourthly, 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.   If a study is pursued on this issue, then applicable use cases, implementation complexity and specification effort must be considered firstly **(low)Potential proposal 6.1A:**  For BM-Case2 (both UE-sided and NW-sided model), for studying on whether~~/how~~ to extend the Rel-17 TCI state activation/indication ignaling methods to activate/indicate N [joint] TCI states which are corresponding to N future time instances, study from the following aspects:   * Potential benefits * Applicable cases, e.g. whether it is applicable or beneficial to K>1, when PDSCH is scheduled, limited by legacy max number of activated TCI states * Implementation complexity and RAN 4 impact * ~~FFS: maximum number for N~~ * ~~FFS: Time periods that each indicated TCI state is applicable.~~   Note: Legacy TCI framework can already support the indication of TCI states according to inference outcome |
| TCL | The TCI state IDs may be extended to support new TCI states specific for AI/ML BM, this approach minimizes the impact to legacy UE. It is also possible to enable one new TCI state contain time information associated with multiple future time instances. |
| NTT DOCOMO | Support. |
| CATT | We prefer not to extent rel-17TCI. 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. |
| NEC | We think the extension is based on introducing the timing information for the indicated TCI states. |
| Ericsson | Not support.  Don’t see a motivation on indicating to the UE the N such TCI states. Based on:   1. This assumes there is no Top-K beam sweep (is unlikely) 2. The overhead saving is questionable   Regarding the beam indication, the main issue is how to indicate the correct TCI states for the Top-K beam sweep at the UE. Activating TCI states for the UE data transmission can be done dynamically using legacy.  However this dynamic method is not available for the Top-K measurements. Hence, our proposal is to take a step back and discuss how the UE can understand which beam that are part of top-k sweep. |
| SPRD | We think there is no need to extent Rel-17 TCI framework. |
| LG | Support to study. This can achieve overhead reduction on beam indication in case of HST scenarios. |
| Fujitsu | Generally ok with the proposal. |
| Google | Support |
| CMCC | Support FL’s proposal. We also accept to further study and discuss the potential indication of multiple TCI states corresponding to the N time instance, which seems reduce the overhead of TCI indication and corresponding PDCCH. |

#### Issue 2: 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

##### **(low)Questions to answer**

A. Whether UE needs to assume RS resources for Set A are always available and/or have been measured?

- if yes, no issue (No TCI without RS)

- if no, any issue

B. What else needs to be studied?

|  |  |
| --- | --- |
| Company | Comments |
| FL | Pls provide some answers |
| TCL | A: Yes. |
| Vivo | At least the Top K beam of Set A can be assumed as available and/or have been measured? |
| CATT | If the beam in Set A is not measured, the indicated TCI state may be unknown TCI state, and the legacy requirements can be reused. |
| NEC | We don’t understand the logic for Question A.  UE needs to assume RS resources for indicated TCI states are always available and/or have been measured, but RS resources for Set A are NOT always available, that is the issue to be solved, why saying “- if yes, no issue”? |
| LG | “NO”, it can be unknown TCI from UE perspective. However, If there is assistance information on relation/association between Set A beams and Set B beams, e.g., indicating multiple neighboring beams from Set B for helping UE to find its Rx beam for the Set A beam, the answer can be “Yes”. |
| Fujitsu | Open for discussion on this issue. |
| Google | Yes after the TCI is activated, no before the TCI activation |
| CMCC | A: NO. we cannot assume the RS resource for set A are always available if the UE does not have any configuration information of the RS resource of the set A.  Whether UE can determine or assume the RS resources for Set A available or measured, depends on the configuration of gNB. If the set A is configured for UE to measure, the UE will measure the configured RS. But if the gNB does not have any configuration of any RS resources in the set A for the measurement, there is no need for UE to take the measurements and event that the UE has no knowledge of the existence of the reference RS.  For the usage of the TCI states, the reference RS of the TCI state should be configured to the UE and the UE has taken the measurements or the reception of the reference RS for determination of the TCI states. |

## Configuration for NW sided model

|  |
| --- |
| 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 contributions:

|  |  |
| --- | --- |
| Company | Proposals |
| HW/HiSi[3] | For NW-sided model training/monitoring, 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., 256.  • Note: Purpose, such as above “For NW-sided model training/monitoring”, will not be specified in RAN 1. |
| Intel [4] | Proposal 1: For a network-side AI/ML model, for BM-Case 1/2, implicit configuration of set B for AI/ML model input can be used based on existing CSI-ResourceConfig and CSI-ReportConfig frameworks. Configuration of set A is not required.  Proposal 6: 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.  Proposal 7: 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.  For a network-side AI/ML model, for BM-Case 2, the UE may not need to be configured with a prediction window. |
| Vivo [9] | Proposal 34: For model inference with NW-side model, support to configure multiple Set B patterns in set B configuration. |
| 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 RS configuration enhancement at least for the case that Set B and Set A are the same:**   * **Configure one RS set for measurement of Set B beams within the measurement window;** * **Configure multiple RS sets for measurement of Set B beams within the measurement window.** |
| CMCC[14] | **Proposal 2: Regarding data collection for NW-side model, following options are considered for the configuration of Set A:**   * **Option 1: UE capability on maximum number of RS for RSRP measurement per resource set is enlarged** * **Option 2: more than one resource set in one *CSI-ResourceConfig* is supported**   **Proposal 7: Regarding configuration of pre-configured set B pattern, following options are considered:**   * **Option 1: configure union of set B patterns with one resource set, configure different set B patterns with bitmap** * **Option 2: configure multiple associated *CSI-ReportConfig ,* each *CSI-ReportConfig* includesone resource set corresponding to one set B pattern** * **Option 3: configure one *CSI-ReportConfig* with multiple *subCSI-ReportConfig ,* each *subCSI-ReportConfig* includesone resource set corresponding to one set B pattern** |
| Lenovo [16] | **Proposal 15: 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.** |
| Fujitsu [20] | ***Proposal 6:***   * *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 7:***   * *RAN1 to further discuss whether the same or separate reference signal configuration among training data collection, model inference and performance monitoring could be applied.* |
| Xiaomi [21] | ***Proposal 3-6: 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.*** |
| ZTE [24] | ***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.*** |

#### Issue #1: Whether to configure multiple resource sets associated to one L1 beam report for NW sided model

##### **(low)Questions to answer**

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.

|  |  |
| --- | --- |
| Company | Proposals |
| TCL | Agree with the first bullet.  For the second bullet, the benefit of time stamp information is not that clear to confirm the bullet, and it may be covered by dataset ID or associated ID (if applicable). |
| Fujitsu | Is this proposal for inference or monitoring or training data collection? |
| Google | We think this depends on the use case. For data collection for training, multiple sets can be considered. |

## 8 Consistency for additional condition

|  |
| --- |
| 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. |

Summary of the discussions in contributions.

|  |  |
| --- | --- |
| Company | Proposals |
| Ericsson [2] | Proposal 1 For UE-sided models, for the consistency of NW-side additional condition across training and inference, further study how/where the additional ID is introduced, using the following alternatives as a starting point: *🡺 FL: CSI framework*  • Identifier providing consistency in CSI-ReportConfig  • Identifier providing consistency in CSI-ResourceConfig  • Identifier providing consistency in CSI-ResourceSet  • Identifier providing consistency in CSI-Resource  • Other alternatives are not precluded  • Note: Consistency in terms of the NW transmission parameters for each alternative  Proposal 2 For UE-sided models, when UE receives the same associated ID, it can assume that NW transmission properties of set A/B are consistent from training to inference,  • Note: Further clarify the definition when it is agreed where/how the identifier is introduced.  Proposal 3 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 |
| Huawei/HiSi[3] | *Observation 4: For the data collection for the UE-side model, the impact factors for NW-side additional condition in the same cell, if any, are with limited variations and low frequent changes.*  *Observation 5: 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 14: For the consistency of NW-side additional condition across training and inference for UE-sided model, study associated ID subject to cell specific manner as a starting point.*   * *If needed, consider to indicate the associated ID by reusing the CSI framework.*   *Proposal 15: For the consistency of NW-side additional condition across training and inference for UE-sided model, UE side performance monitoring can also be considered.* |
| Intel [4] | Proposal 25: 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 assumptions for resources configured in the set are consistent when the same ID is used.  Proposal 26: For a UE sided AI/ML model, for consistency between training and inference, performance monitoring-based approaches should be deprioritized. |
| H3C [5] | For the UE side model, if Set B is not a subset of Set A, consider the following two methods to ensure a consistent understanding of the relationship between Set A and Set B during the training and inference stages:  • Opt1: The mapping relationship between beams in Set A and Set B remains unchanged during both training and inference stages.  • Opt2: The mapping relationship between Set A and Set B can change, but follows the same physical beam pointing order, such as horizontal and vertical direction. |
| Spreadtrum [7] | *Proposal 14: For inference for UE-side models, to ensure consistency between training and inference, option 1 should be considered.* |
| Samsung [8] | **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* for inference results reporting.**   **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,**   * **Further study the method to ensure the consistency of downlink spatial domain transmission filters corresponding to the beams in Set A and Set B.** |
| Vivo [9] | **Proposal 5: Associated ID together with RS/report configuration can be used to implicit indicate purpose of resource configuration for Set B and/or Set A, e.g., whether it is for data collection for UE sided model or not.**  Due to Set B and/or Set A can be linked to both of a associated ID and a beam report configuration with report quantity = none, UE can assume that this Set B and/or Set A are configured for data collection purpose.  **Proposal 17: For model inference with 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.**  **Proposal 18: Based on Rel-18 study, the same associated ID can be assumed to imply the same network-side additional conditions including the same Tx beam boresight direction (azimuth and elevation), the same 3dB beam-width and the same mapping of actual beams to beam indices.**  ** It can be further studied how to capture such assumption into specification.**  **Proposal 19: Global associated ID can be optionally supported.**  **Proposal 20: Local associated ID can be supported with the understanding that the model is managed in a cell/site/region specific way.** |
| Apple [10] | **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 assistance information/associated ID (e.g., dataset ID/model ID), if assigned by higher layer is embedded as part of reference signal configuration.** |
| Interdigital [11] | ***Proposal 16:*** *Support both Opt.1 (based on an associated ID) and Opt.2 (performance monitoring based) for the consistency of NW-side additional conditions.* |
| CATT [12] | **Proposal 26: The following additional conditions could impact the AI/ML model performance if they are not consistent between training and inference:**   * **Tx beam codebook;** * **Association of Set B and Set A;** * **The order of model input and model output.**   **Proposal 27: 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 should be discussed.**  **Proposal 28: 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 29: 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.**  **Proposal 30: To ensure the consistency of patterns/association of Set B and Set A, the QCL relation of RS with Set B beams and Set A beams can be considered.**  **Proposal 31: To ensure the consistency of model input/output order, the mapping rule between RS and Tx beams can be pre-defined.** |
| China Telecom [13] | ***Proposal 8. For UE-side AI/ML model, consider the use of data collection ID/dataset categorization information across LCM operation to ensure consistency between training and inference.*** |
| CMCC [14] | **Proposal 12: 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)**   + **At least beam number of Set A/Set B and association of Set A/Set B can be assumed by UE with the same associated ID across training and inference *=>FL: This can be treated as condition other than additional condition***   + **associated ID is introduced within CSI framework** * **Opt 2: Performance monitoring based** |
| NVIDIA [17] | **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.** |
| LGE [18] | **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** |
| Fujistu [20] | ***Proposal 34:***   * *Regarding the consistency across training and inference, the option based on performance monitoring is preferred.* |
| Xiaomi [21] | ***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: The following NW side additional conditions need to be indicated to UE explicitly to ensure consistency between training and inference for UE side model.***   * + Set B/ set A configuration *=>FL: the following part seems possible to be described by signaling. Therefore, no need to define this as “additional condition” but condition*      - The number of beams in set B     - The number 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) *=>FL: the following part seems possible to be described by signaling. Therefore, no need to define this as “additional condition” but condition*     - The number of history measurement time instance     - The number of predicted future time instance   + Deployment     - Transmission power and antenna height   + UE distribution *=>FL: for UE sided model, or for a given model, UE distribution is not an issue. This cannot be controlled either by gNB/NW/Operator/Ues. And in SI phase, this doesn’t impact the performance significantly.*   The second type is the additional conditions which are the sensitive proprietary information and NW operators may have some concerns to indicate them to UE. Such as the following information. One alternative way to ensure the consistency of such kinds of information is to indicate the relative information, or indicate the virtual ID of each configuration. It means that NW side maintains the mapping between the virtual ID and the configuration, but UE just need to know the virtual ID to select the most suitable model.  - NW-side beam shape information   * E.g., 3dB beamwidth, beam boresight directions, beam shape, Tx beam angle, etc. |
| NEC[22] | Proposal 3: Support associated ID 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.  − Associated ID shall at least indicate the site/cell specific variables used for model training like antenna configuration.  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 association between CSI-RS in AI/ML model training phase, CSI-RS in AI/ML model inference phase, as well as CSI-RS in AI/ML performance monitoring phase.***  Proposal 8: For avoiding the proprietary/privacy issue, study to provide the assistance information (e.g., angle related information, channel estimation based information) implicitly from one side to the other side. |
| ZTE [24] | Proposal 28: RAN1 concludes concrete NW-side additional conditions specific to AI/ML beam management and then discuss the potential consistency approaches if necessary, taking the following as baseline  • Deployment scenarios (e.g., ISD, Umi/Uma), gNB antenna array dimensions, DL Tx beam codebooks, and Set A/B patterns (e.g., indexing/mapping between Set A and Set B).  Observation 1: Associated ID based approach to align NW-side additional conditions across training and inference can be challenging due to the following concerns.  • Matching ID granularity and model capability  • Signaling overhead for ID exchange  • Proprietary information disclosure risks  Observation 2: Performance monitoring-based approach can be useful for the alignment of NW-side additional conditions and communication environment changes with minor spec effort foreseen. |
| ETRI [27] | **Proposal 14: For UE-sided model, following procedure can be considered as a solution for resolving consistency between training and inference**   * **A: For data collection, the NW signals the data collection related configuration(s) and its/their associated ID(s)** * **B: UE(s) collects the data corresponding to the associated ID(s)** * **C: AI/ML models are developed (e.g., trained, updated) at the UE-sided based on the collected data corresponding to the associated ID(s).**   **Proposal 15: For AI/ML beam management, support is provided for restricting the beam patterns and indexing scheme to ensure consistency between training and inference across multiple cells.** |
| OPPO [29] | Proposal 23: To ensure consistency between Set A and Set B across training and inference, adopt (Opt1) the associated ID (from A.I. 9.1.3.3) as a proper solution. |
| Fraunhofer [30] | **Proposal 13: Functionality LCM with model-ID or model-ID LCM should be supported for UE-sided models, at least for BM-Case 1.**  **Proposal 14: The UE capabilities report may indicate its supporting condition, to allow configuration of functionality LCM or model-ID LCM.**  **Proposal 15: For UE-side models, support signaling of assistance information from the NW to the UE, at least for BM-Case 1.** |
| Nokia [31] | **Proposal 12: For beam prediction use-cases, RAN1 shall support a solution to ensure consistency between training and inference regarding NW-side additional conditions.**  **Proposal 13. 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.**   * **The purpose of the identifier is to differentiate or relate NW-side assumptions used in the CSI-RS or SSB transmissions associated with different CSI-RS resource configurations.** * **Assigning identifiers (associated ID) shall be left to NW-implementations.**   **Proposal 14: 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 ignaling support.** |
| DoCoMo [32] | **Proposal 3: Support configuring associated ID within resource related configuration for Set A/B, where the corresponding Set A/B can be assumed to be consistent for the same associated ID.**  **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 [37] | **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 condition across training and inference, with regards to FFS on what can be assumed by UE with the same associated ID across training and inference:**  **• 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.**  **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 at least Opt1: Based on associated ID.**  **Proposal 4**  **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.** |
| KT [35] | ***Proposal 2. Prefer to defer the discussion on model/associated ID-based option until the study on Model ID in AI 9.1.3.3 is completed.***  ***Proposal 3. Discuss the details for the performance monitoring-based as a mechanism to ensure consistency between training and inference for UE-side models.*** |

#### Issue #1: How to configure the identifier for UE sided model

Associated ID is configured within CSI framework

* Supported by: (6)Ericsson, Samsung (in *CSI-ReportConfig*), Apple (Reference signal conf), CMCC, xiaomi, Nokia(CSI-RS resource configuration)

Cell specific

* Huawei, vivo

Global

* Vivo (optionally), apple? (PLMN unique)

#### Issue #2: UE assumption with the identifier for UE sided model

UE assumptions with the identifier:

* QCL assumption
  + Supported by(2) Intel, CATT
* Mapping relationship of Set A and Set B, including ordering to (a set of ID, or resource )
  + Supported by (6): H3C, CATT, CMCC, ZTE, DoCoMo, Qualcomm
* **Consistency of downlink spatial domain transmission filters corresponding to the beams in Set A and Set B.**
  + **~~the same Tx beam boresight direction (azimuth and elevation), the same 3dB beam-width and the same mapping of actual beams to beam indices.~~**
  + ~~NW transmission properties of set A/B are consistent from training to inference~~
  + ~~E.g., 3dB beamwidth, beam boresight directions, beam shape, Tx beam angle, etc.~~
  + ~~gNB antenna array dimensions, DL Tx beam codebooks~~
  + **~~Beam shape consistency:~~**
  + Supported by (8): Ericsson, Samsung, vivo, CATT, LGE, xiaomi, ZTE, Qualcomm
* **The order of model input and model output.**
  + CATT
* **between RS and Tx beams can be pre-defined.**
  + **CATT**
* **Transmission power** 
  + **Xiaomi**
* UE distribution
  + Xiaomi
* **antenna height**
* Deployment scenarios (e.g., ISD, Umi/Uma)
  + ZTE
* **ensure consistency across different cells.**
  + **Qualcomm**

#### Issue #3: NW-sided consistency

|  |  |
| --- | --- |
| Company | Proposal |
| Ericsson [2] | Proposal 25 For a NW-sided model, to ensure consistency from training to inference, study the feasibility of the following mechanisms,  • UE indicate RxBeamIndex during set A/B data collection.  • NW indication that UE could use fixed RX-beam during set A/B data collection.  • UE indicates its RSRP measurement accuracy |
| Huawei/HiSi[3] | Observation 3: Regarding data collection for NW-side AI/ML model, the legacy QCL mechanism can be reused to guide the UE for its Rx beam selection. |
| Intel [4] | Proposal 10: RAN1 should further discuss if and how the same UE Rx beam assumption can be maintained for measurement of set A for data collection for model training and for measurement of set B for model input of the corresponding model.  Proposal 24: 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. |
| Vivo[9] | Proposal 8: For data collection procedure with 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. |
| CATT [12] | **Proposal 17: For NW-sided model, it is beneficial to align the Rx information of the measurements between network and UE.** |
| CMCC[14] | **Proposal 3: Rx beam assumption for a measurement report can be up to gNB implementation.** |
| Fujitsu [20] | ***Proposal 9:***   * *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).* |
| Xiaomi [21] | ***Proposal 3-5: Exchange the UE-side additional condition such as Rx beam assumption and UE speed during the procedure of data collection for NW-side AI/ML model training.*** |
| OPPO [29] | ***Proposal 8: For BM-Case1 and BM-Case2 with NW-side model, do NOT specify additional condition on UE Rx beam assumption.*** |
| DoCoMo [32] | **Proposal 18: If NW side beam prediction gets difficult due to UE side additional condition (e.g., UE Rx beam assumption), some enhancements should be introduced.** |
| Nokia [31] | * **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.** |

### 8.1 1st Round discussion

#### Issue 1: Associated ID for UE sided model

##### **(high)Proposal 8.1A:**

Support associated ID to ensure 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,

* The associated ID is configured within CSI framework (with RS resource configuration), FFS on details
* FFS on whether performance monitoring/validation for model activation

|  |  |
| --- | --- |
| Company | Comments |
| FL | I don’t see any complain for supporting this.  All companies mentioned about configuration, support to use CSI framework.  In addition, as commented in performance monitoring, I have a question on whether a model can be directly used (activated) as long as the associated ID are the same, without performance checking? Or, in some situation, the performance should be checked first (e.g., 3 months later, there may not have leaves on the tree so that the environment changed even with the associated ID). This is the motivation of the last FFS. |
| HW/HiSi | For the identifier we first study its range, whether it is applicable on cell level or larger. In our view it should be cell specific rather than a globally unique ID or area unique ID, i.e., the ID can be used to indicate the variation of the additional condition of a single gNB over time, but should not disclose the linkage across gNBs. From our paper review, we concluded that several companies want to at least start we a cell specific ID.   **High)Proposal 8.1A:**  For ~~Support~~ associated ID to ensure 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, study the applicable range of the ID, with cell specific identifier as starting point.   * ~~The associated ID is configured within CSI framework (with RS resource configuration), FFS on details~~   ~~FFS on whether performance monitoring/validation for model activation~~ |
| TCL | Support the 1st bullet.  The FFS depends on the exactly definition of model ID, consequently the relationship between model ID and associated ID. We suggest to postpone the discussion of the FFS here until 9.1.3.3 gives a clear answer on the two questions. |
| Vivo | Support the main bullet. The whole subbullet can be in FFS. We don’t think it’s the right time to discuss these details for now. |
| NTT DOCOMO | Support the principle of proposal. In addition, we think it is better to study the relationship between associated ID and applicability reporting. With that in mind, we suggest the following modification for FFS.  FFS on whether/how applicability reporting/performance monitoring/validation for model activation |
| QC | Support FL’s proposal. |
| CATT | ok |
| ETRI | As FL’s comment, performance monitoring may be needed before model deployment or activation. But I think the relationship between Associated ID and the AI/ML model (model ID) is unclear. So, performance monitoring with Associated ID is unclear for now. |
| ZTE | At the current stage, the scope of the associated ID is not clear, e.g., whether it is a local or region-specific ID to be applied within a local area, or it is a global ID. Besides, there are still some feasibility issues related to the associated ID assignment especially if the associated ID is global, such as matching ID granularity and model capability, signaling overhead for ID exchange, and proprietary information disclosure risks. Therefore, more clarifications are needed before the supportive of associated ID or not.  Regarding the FFS, as model transfer or pairing is not considered for AI BM, the functionality-based LCM without specifying any model ID based signaling should be utilized and the model-level LCM should be transparent to the NW.  Accordingly, we suggest the following revisions.  ~~Support~~ For the study of associated ID to ensure 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,   * The associated ID is configured within CSI framework (with RS resource configuration), FFS on details * FFS on whether performance monitoring/validation for ~~model~~ functionality activation |
| Panasonic | We support it. |
| Xiaomi | Support the main bullet and the first sub-bullet, open to discuss the last FFS. |
| Intel | Support in principle, but the first sub-bullet is premature.  We do not need to commit to tagging along with CSI config along with resources already. It depends on resolution of how to configure set A of beams. |
| NEC | OK to the principle to use associated ID to ensure the consistency, but we still do not know what the “associated ID” is. |
| New H3C | OK |
| Ericsson | Support |
| SPRD | Support. |
| LG | Support in principle. Where the association ID is configured may be impacted by the other discussion on Set A configuration for different purposes(e.g. Issue#1 in Section 3). Thus, we suggest to revise the proposal as follows:  **Proposal 8.1A:**  Support associated ID to ensure 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,   * ~~The associated ID is configured within CSI framework (with RS resource configuration),~~ FFS on configuration details of the association ID * ~~FFS on whether performance monitoring/validation for model activation~~ |
| Fujitsu | Don’t support. We think the consistency could be via performance monitoring.  With associated ID, what’s the UE behavior if the associated doesn’t match? |
| Google | Support |
| CMCC | Support in principle. But in our view, the association id is more like the configuration id. For the other cases, it is still not clear on how to define the association id, if the details are not clarified. |
| CAICT | Support. |
| OPPO | Support the FL proposal. |

#### Issue #2: UE assumption with the identifier for UE sided model

UE assumptions with the identifier:

* Consistency of downlink spatial domain transmission filters corresponding to the beams in Set A and Set B.
  + ~~the same Tx beam boresight direction (azimuth and elevation), the same 3dB beam-width and the same mapping of actual beams to beam indices.~~
  + ~~NW transmission properties of set A/B are consistent from training to inference~~
  + ~~E.g., 3dB beamwidth, beam boresight directions, beam shape, Tx beam angle, etc.~~
  + ~~gNB antenna array dimensions, DL Tx beam codebooks~~
  + ~~Beam shape consistency:~~
  + Supported by(8): Ericsson, Samsung, vivo, CATT, LGE, xiaomi, ZTE, Qualcomm
* Mapping relationship of Set A and Set B, including ordering to (a set of ID, or resource )
  + Supported by (6): H3C, CATT, CMCC, ZTE, DoCoMo, Qualcomm => Covered by FFS
* QCL assumption
  + Intel, CATT => see the question
* The order of model input and model output.
  + CATT => Covered by FFS
* between RS and Tx beams can be pre-defined.
  + CATT => Covered by FFS
* Transmission power
  + Xiaomi => see the question
* UE distribution
  + Xiaomi => As previous comment, this is out of control. And not related to UE sided model
* antenna height
  + Xiaomi => see the question
* Deployment scenarios (e.g., ISD, Umi/Uma)
  + ZTE => see the question
* ensure consistency across different cells.
  + Qualcomm=> see the question

##### **(high)Proposal 8.2A:**

For UE sided model, with the same associated ID across training and inference, UE assumes

* UE assume the same DL TX spatial filters of the corresponding beams in Set A and Set B across training and inference
  + FFS on how to determinate/configure the order or index of the corresponding beams within a set (i.e., Set A and/or Set B)
* FFS on other assumptions

|  |  |
| --- | --- |
| Company | Comments |
| FL | The first FFS is related to the “mapping”. And ordering (not the input/output order, but the order/index when UE collect data or obtain the input.)  In my reading and study of the submitted contributions. There are two ways:   1. implicitly, the order of (resources) for the beam in Set A and Set B are kept the same. 2. Introduce an beam ID, with the same beam ID, the tx filter is the same. This ID can be the CRI or SSBRI of Set A. and Set B.   For other assumptions.   * **QCL:** for UE to terminate Rx filter, I think if the Tx filter is the same, no need to mention QCL. But with QCL, it may not be strong enough * **Tx power:** I am not sure. Maybe the difference between beams in a set can be consistency. But, no need to ensure the Tx power to be always the same. Another question, whether this is covered by Tx spatial filter? * **Antenna height and down tilt:** based on the simulation. This will impact on the performance. For cell specific level, maybe OK. Not sure for Global * **Deployment scenarios:** based on the simulation, without changing Antenna height and down tilt, no much impact. * **Ensure consistency across different cells:** this is a good question. For global ID, it may be possible. But how NW can assign the same ID for different cell? For different cell, the antenna height down tilt, and surrounding most likely to be difference. In addition, as mentioned by some other companies, maybe, for some small varies of some additional conditions, it can be handled by AI model.   Let’s do it step by step, we will discuss all the necessary assumptions. Please don’t block the one has a lot of supports and makes sense, even you want something else. |
| HW/HiSi | An identifier, if introduced, should be limited to cell level to not disclose NW proprietary characteristics. Also, it will be complicated to categorize all globally different aspects into different identifiers.  It is not really clear how to limit specific conditions, such as spatial filter, there are many other conditions that also could change (e.g. power. To reflect the FL intention above, we suggest a more generic wording at this stage.  We are therefore suggesting the following updated proposal:  For UE sided model, with the same associated ID across training and inference, if supported, UE assumes   * The associated ID can be interpreted to indicate an individual sort of channel status feature from NW perspective. * ~~UE assume the same DL TX spatial filters of the corresponding beams in Set A and Set B across training and inference~~   + ~~FFS on how to determinate/configure the order or index of the corresponding beams within a set (i.e., Set A and/or Set B)~~ * FFS on other assumptions |
| TCL | Since the associated ID indicates the NW side additional conditions, therefore it should be a local ID, or even temporal. Since it is introduced to support consistency, it should reflects more additional conditions besides the Tx/Rx beam assumptions. |
| Vivo | support |
| QC | support |
| CATT | Ok |
| ZTE | The different LCM procedures (i.e., training and inference) may be transparent to the NW. From the UE perspective, upon repeatedly receiving the same associated ID at different times, the UE can assume that the same NW-side additional condition (such as a specific beam shape or spatial transmission filter) is being utilized by the network. Therefore, we suggest the following revision.  For UE sided model, with repeatedly reception of the same associated ID at different times ~~across training and inference~~, if supported, UE assumes   * UE assume the same DL TX spatial filters of the corresponding beams in Set A and Set B at different times ~~across training and inference~~   + FFS on how to determinate/configure the order or index of the corresponding beams within a set (i.e., Set A and/or Set B) * FFS on other assumptions |
| Xiaomi | Support and suggest to add ‘number’ in addition to order or index   * UE assume the same DL TX spatial filters of the corresponding beams in Set A and Set B across training and inference   + FFS on how to determinate/configure the number or order or index of the corresponding beams within a set (i.e., Set A and/or Set B) |
| Intel | For similar reasons as given by FL against consideration of Tx power (which we agree with), we think QCL assumption is sufficient and assumption on DL Tx spatial filter is “too strong”. There is no need for NW to commit to no changes as part of the DL Tx spatial filter if some minor changes may still result in a beam of very similar shape as far as TCI identification (corresponding to beam pair link formation) is consistent.  Thus, we prefer to prioritize QCL over Tx spatial filter assumption. Further, we need to first discuss what exactly is involved as part of “DL Tx spatial filter” and if that can be visible in specs. |
| NEC | Not support, suppose the training is via gNB 1, inference is for gNB 2, it is not possible to say DL TX spatial filters are the same. |
| New H3C | OK |
| Ericsson | Support.  The FFS in the sub-bullet is not needed. If the ID is introduced on a CSI-Resource level, there is no need for such ordering. More specifically:   * ***Identifier defined on a Resource-level*** : UE can assume the *NZP-CSI-RS-*resource is transmitted using the same NW transmission properties (e.g. spatial TX-filter) across training and inference * ***Identifier defined on a ResourceSet*-level**: UE can assume that the order/indexing of the resources within the set are consistent, and each respective resource are transmitted with the same NW transmission properties (e.g. spatial TX-filters) across training and inference. |
| LG | Support in principle. We suggest to revise as below, with more specification wording **Proposal 8.2A:**   * Under same association ID for different resources, UE assume the same DL TX spatial filter of the corresponding resources * Under an association ID, UE assume the same DL TX spatial filter of the different transmission instances of the corresponding resource(s) |
| Fujitsu | This proposal could be postponed after decision on which option is used for consistency. |
| Google | OK |
| CMCC | Though we support to use the association id to ensure the consistency between inference and training. But the DL Tx spatial filter is a strong limitation that, the exactly the precoding or weights of the DL beam forming at gNB side should be same. If this is the understanding, UE should first acquire the DL beamforming weights of the D, which is not practical. In the procedure of the inference, the DL beamforming weights depends on the gNB design which may be also proprietary. |
| CAICT | Support. |

### 8.2 2nd Round discussion

#### Issue 1: Associated ID for UE sided model

##### **(high)Proposal 8.1B:**

If associated ID is supported, the associated ID is configured within CSI framework (with RS resource configuration), FFS on details

* FFS on whether performance monitoring/ how applicability reporting /validation for functionality activation

|  |  |
| --- | --- |
| Company | Comments |
| FL | Let’s see whether this can work or not |
| OPPO | It seems proper to configure the associated ID within CSI framework where Set B and/or Set A is configured.  But regarding the FFS, we don’t quite understand the meaning of “applicability reporting/validation for functionality activation”. It sounds too vague, can the proponent(s) elaborate more on it? |
| MediaTek | Does this proposal mean associated ID is configured/represented in the form of RS resource configuration? It seems like this proposal excludes the option that associated ID can be aligned between NW and UE through LCM-related signaling (i.e., when AI model is activated/switched)? |
| Hw/HiSi | The direct of the main bullet seems fine, but we would like to emphasize that a cell-level ID should be the baseline, or at least the starting point of the discussion.  Regarding the FFS bullet we share the comment from Oppo and also think that this would be a separate discussion.  Potential update:  *If associated ID is supported, the associated ID is per cell level and configured within CSI framework (with RS resource configuration), FFS on details.* |
| Intel | Do not support. We do not need to decide on where the associated ID is configured – especially if we still cannot agree on supporting associated ID. As commented in the 1st round, this can follow once details of configuration of sets A and B achieve further clarity. |
| ZTE | Share similar view with Intel. If the scope of the associated ID and the Set A/B configuration are not clear, it would be too early to decide where to configure the associated ID. Additionally, if the associated ID is to abstract the NW-side additional conditions, it shall be intended to be shared by multiple UEs by cell-specific signaling instead of being configured by UE-specific signaling. |
| InterDigital | Support |
| CEWiT | Support |
| CATT | Do not support. Before discussing how to configure associated ID, we should make it clear what is associated ID and how it can work across training and inference. And it still under discussion on the relationship of model id and associated id. If one associated ID is related to multiple model id, we are not sure the associated ID is configured in CSI framework.  The associated ID represents additional condition of network, it can be used to provide some undefined network information, it may be not associated with a certain CSI configuration. |
| NEC | We would like to understand why it is configured with RS resource configuration, but not with others, e.g., report configuration, in the CSI framework. |

#### Issue #2: UE assumption with the identifier for UE sided model

For UE sided model, with the same associated ID across training and inference, UE assumes

* The consistency of the order of resources (corresponding to beams) for Set A of beams across training and inference
* The consistency of the order of resources (corresponding to beams) for Set B of beams across training and inference
* FFS on the details including, whether to introducing beam ID, or whether a virtual resource or no resource can be configured to a beams

|  |  |
| --- | --- |
| Company | Comments |
| FL | Tx filter may be too aggressive for this meeting. How about the ordering? |
| OPPO | Support the ordering, otherwise we don’t know to guarantee the consistency. |
| MediaTek | We are generally fine with the direction. However, with the same associated ID, UE should also assume the type of Set B beam is the same (UE may not need to know exactly its SSB or CSI-RS, but at least UE should assume the beam type/shape is the same), right? Suggest put a “• FFS: other UE assumptions on NW side conditions” |
| HwHiSi | We would like to re-iterate our comment from the first round, rather than listing all possible assumptions (which we could become an endless list) will should combine the assumptions in or more generic term.  Suggested update:  For UE sided model, with the same associated ID, if supported, across training and inference, UE assumes   * ~~The consistency of the order of resources (corresponding to beams) for Set A of beams across training and inference~~ * ~~The consistency of the order of resources (corresponding to beams) for Set B of beams across training and inference~~ * ~~FFS on the details including, whether to introducing beam ID, or whether a virtual resource or no resource can be configured to a beams~~ * The same individual sort of channel status feature from NW perspective. |
| Intel | OK |
| Apple | OK |
| InterDigital | Fine |
| TCL | We suggest to change the main bullet like this:  For UE sided model, with the same associated ID across training and inference, UE at least assumes: ….  Since the associated ID may content other information besides the beam related information. |
| CEWiT | OK |
| CATT | We do not think all the above aspects can be assumed with the identifier, but they may impact the AI/ML model performance if the above consistency is not kept between training and inference. We prefer to first discuss the applicable range of the identifier. |

## 9 Others

#### Issue #1: For UE sided model, AI/ML processing capability

* Huawei/HiSi [3]: Proposal 33: For UE-side model of BM, the legacy mechanism of CSI processing unit and CSI processing timeline can be reused as a starting point to represent the processing capability of AI/ML-based CSI report.
  + As an enhancement, the AI/ML-based CSI processing capability can be reported by UE for per functionality, considering different complexities for different models/functionalities.
* 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 [30] 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.

#### Issue #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 [4] 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 [23] 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.
* OPPO [9] For temporal domain beam prediction, suggest to study and evaluate the beam dwelling time prediction.
* Fujitsu [20] 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 [32] Discuss how to handle measurement sensitivity issue in the measurements of Set B/C.

#### Issue #3: How to define “a condition” for UE sided model

Xiaomi [21]

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.

#### Issue #4: Others

Ericsson [2]

Proposal 10 Further study how Top-K beam measurements 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 to configure CSI-Resources for all possible measurement combinations of Top-K beams

# References

1. R1-2403866 Discussion on specification support for beam management FUTUREWEI
2. R1-2403914 AIML for beam management Ericsson
3. R1-2403929 Discussion on AI/ML for beam management Huawei, HiSilicon
4. R1-2403973 Specification support for AI/ML for beam management Intel Corporation
5. R1-2403998 Discussions on AIML for beam management New H3C Technologies Co., Ltd.
6. R1-2403999 Discussion on AI/ML beam management TCL
7. R1-2404015 Discussion on AIML for beam management Spreadtrum Communications
8. R1-2404137 Discussion for supporting AI/ML based beam management Samsung
9. R1-2404165 Specification support for beam management vivo
10. R1-2404272 Discussion on AI/ML-based beam management Apple
11. R1-2404311 Discussion on AI/ML for beam management InterDigital, Inc.
12. R1-2404384 Discussion on AI/ML for beam management CATT
13. R1-2404421 Discussion on specification support for AI/ML beam management China Telecom
14. R1-2404444 Discussion on specification support for beam management CMCC
15. R1-2404490 Discussions on AI/ML for beam management Sony
16. R1-2404525 AI/ML specification support for beam management Lenovo
17. R1-2404536 Specification support for AI-enabled beam management NVIDIA
18. R1-2404546 Discussions on AI/ML for beam management LG Electronics
19. R1-2404567 Discussion on specification support for beam management Panasonic
20. R1-2404582 Discussion on specification support on AI/ML for beam management Fujitsu
21. R1-2404601 Specification support for beam management Xiaomi
22. R1-2404655 Discussion on specification support for beam management NEC
23. R1-2404682 AI/ML based Beam Management Google
24. R1-2404701 Discussion on specification support for AI/ML beam management ZTE
25. R1-2404721 Discussions on AI/ML for beam management CAICT
26. R1-2404737 Discussion on AI/ML based beam management Hyundai Motor Company
27. R1-2404766 Discussion on specification support for beam management ETRI
28. R1-2404802 Prediction of untransmitted beams in a UE-side AI-ML model Rakuten Mobile, Inc
29. R1-2404877 On specification for AI/ML-based beam management OPPO
30. R1-2404903 Specification support for beam management Fraunhofer HHI, Fraunhofer IIS
31. R1-2404904 AI/ML for Beam Management Nokia
32. R1-2405030 Discussion on AI/ML for beam management NTT DOCOMO, INC.
33. R1-2405068 Discussions on specification support for beam management Sharp
34. R1-2405087 Discussion on specification support for AI/ML-based beam management MediaTek Inc.
35. R1-2405096 Discussion on AI/ML based beam management KT Corp.
36. R1-2405121 Discussions on specification support for beam management Ruijie Networks Co. Ltd
37. R1-2405143 Specification support for AI-ML-based beam management Qualcomm Incorporated
38. R1-2405223 Specification support for AI/ML beam management ITL
39. R1-2405234 Discussion on Specification Support for Beam Management CEWiT
40. R1-2405284 Discussions on Specification Support of AI/ML for Beam Management Indian Institute of Tech (M), IIT Kanpur
41. R1-2405336 Specification support for beam management KDDI Corporation

# Previous agreements

## 8.1 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**

## 8.2 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.