**3GPP TSG RAN WG1 #116-bis R1-2403570**

**Changsha, Hunan Province, China, April 15th – 19th, 2024**

**Agenda item:** 9.1.1

**Source:** Samsung (Moderator)

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

**Document for:** Discussion and Decision

# Introduction

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

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

In this contribution, summarized the contributions in RAN 1 #116 on AI/ML for beam management.

## FL0: Question 0

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

|  |  |  |
| --- | --- | --- |
| Company | Contact | Email address |
| New H3C | Lei Zhou | Zhou.leih@h3c.com |
| Indian Institute of Technology Madras | Sai Prasad Pirati | venkatasiva@5gtbiitm.in |
| OPPO | Jeffrey (Jianfei) Cao | [caojianfei@oppo.com](mailto:caojianfei@oppo.com) |
| Ericsson | Henrik Ryden | Henrik.a.ryden@ericsson.com |
| Fujitsu | WANG Guotong (David) | wangguotong@fujitsu.com |
| MediaTek | Yu-Jen Ku | [yu-jen.ku@mediatek.com](mailto:yu-jen.ku@mediatek.com) |
| TCL | Pu Yuan | pu.yuan@tcl.com |
| CMCC | Yi Zheng  Jiazhen Zhang | zhengyi@chinamobile.com  zhangjiazhen@chinamobile.com |
| NTT DOCOMO | Haruhi Echigo | haruhi.echigo.fw@nttdocomo.com |
| InterDigital | Youngwoo Kwak | [youngwoo.kwak@interdigital.com](mailto:youngwoo.kwak@interdigital.com) |
| vivo | Hao Wu | hao.wu@vivo.com |
| NEC | Peng Guan  Pravjyot Deogun | [Guan\_peng@nec.cn](mailto:Guan_peng@nec.cn)  pravjyot.deogun@EMEA.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 |
| 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 | |

## 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] | * Set A and Set B may or may not need to be differentiated/explicit configured. * Set A and Set B may or may not be configured with association. | Configuration of Set A   * Alt-1. Set A corresponds to RS resources * Set A may or may not need to be measured. * Alt-2. Set A does not correspond to RS resources * [TCI for Set A] |
| Measurement Report contents  (For training/inference/  performance monitoring) | Purpose: training/[inference]/monitoring  Content:   * RSRP of a set of beams (Set A/Set B/Subset of Set A) * Top 1/K index (no for inference) * [Time stamp, multiple instances]   Signalling:   * at least L1, FFS others (for training?) | Purpose: [training]/[inference]/monitoring  Content:   * RSRP of a set of beams (Set A/Set B/Subset of Set A) * Top 1/K index * [Time stamp, multiple instances]   Signalling:  at least L1 |
| Beam indication | TCI indication for BMCase-2 | TCI indication for BMCase-2 |
| Consistency and additional condition | FFS   * Rx assumption | * How to ensure the consistency/handle additional information |
| Performance monitoring  (Other than measurement report) | Metrics and Procedures | Metrics and Procedures |

# Resource set configuration

|  |
| --- |
| **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) |

## 2.1 For NW-sided model:

**FL0: Proposal 2.1b (Config for Set A and Set B)**

For network-sided AI/ML model for BM-Case1 and BM-Case2 for training, inference, and monitoring, support using existing CSI framework to configure resources for Set A and Set B beams as the starting point.

* FFS on necessary enhancement, including whether/how a measurement window or a number of measurement instances can be configured with the measurement resource set for BM-Case2.
* Note: Purpose, such as above “For NW-sided model, for BM-Case1 and BM-Case2 for training, inference, and monitoring” and “Set A and Set B beams”, will not be specified in RAN 1 specifications

**FL1: Proposal 2.1c (Config for Set A and Set B)**

For network-sided AI/ML model for BM-Case1 and BM-Case2, support using existing CSI framework to configure resources for Set A and Set B beams as the starting point.

* FFS on necessary enhancements, including
  + Enhancement for BM-Case2
  + For Set A configuration,
    - Alt 1: whether to enlarge of the max number of resources per resource set (i.e., maxNrofNZP-CSI-RS-ResourcesPerSet)
    - Alt 2: whether to support more than one resource set in in one *CSI-ResourceConfig*
  + whether to separately configure for Set A and Set B in one CSI-ReportConfig
  + whether to separately configure two resource sets for Set A and Set B in CSI-ResourceConfig
* Note: Purpose, such as above “For NW-sided model, for BM-Case1 and BM-Case2 for training, inference, and monitoring” and “Set A and Set B beams”, will not be specified in RAN 1 specifications

|  |  |
| --- | --- |
| Company | Proposals |
| FL | Major companies propose to use existing CSI framework for NW sided model. Details are further studied.  Reports will be separately discussed later. |
| OPPO | Could we ask the purpose for configuring Set A and Set B? Since the intention is not clearly mentioned, we understand the configurations are applicable for training, inference, and monitoring. If that’s the case, configurations can be different. For NW-side model inference (UE only reporting Set B), it seems not necessary to configure Set A. But for model training and monitoring, configuring Set A has some good reasons to do so. |
| Xiaomi | We support to use existing CSI framework to configure resources for Set A and Set B. but we suggest to add one more FFS.  FFS: whether to use one or two CSI-ReportConfig/ CSI-ResourceConfig/ resource set  In addition, as for the second note, could FL clarify the use case of define/configure “Set A and Set B beams” explicitly?  FL: I thought about it. But different form UE-sided model, for NW sided model, there is no motivation to change current link. |
| Fujitsu | Generally fine with FL proposal.  But for the following bullet, we think the CSI-ResourceConfig is associated with one or multiple resource sets instead of resources. If the intention is to extend the max number of resources, then the wording “enlarge of the max number of resources in one CSI-RS resource set” might be more accurate.   * FFS on whether to enlarge of the max number of resources in one *CSI-ResourceConfig* |
| MediaTek | We prefer to remove the second FFS. Note that for NW-side data collection for training, NW needs to configure Set A (with the enlarged max number of resources) for UE to measure. The corresponding UE side measurement behavior is the same as legacy beam management, where UE measures all the Set A of beams. Therefore, if we want to enlarge the max number of resources in a CSI-ResourceConfig, we need to guarantee that the legacy beam management can support this new enlarged max number for UE to measure too. Therefore, we suggest discussing this in the main MIMO session. |
| ZTE | For the second subbullet, enlarging the maximum number of resources within a CSI-ResourceConfig does not have a direct correlation with enabling AI/ML features. Instead, the crucial factor influencing beam prediction performance is the ratio of beams in Set B to those in Set A, rather than solely the number of beams in Set A. Additionally, as the second subbullet pertains to the discussion of UE capability, it can be deferred to a later stage. |
| HW/HISi | 1. For the “Set A and Set B beams”, we suggest to using the same description of the agreement in 116 meeting for inference, i.e., the purpose, such as above “Set A and Set B beams” will not be specified in RAN 1 specifications – RAN1 does not need to specify the purpose of CSI measurement is for obtaining label or inference input by gNB.  2. For the second FFS, we think the intention is to enable UE to measure larger number of beams (for Set A) under one CSI report. Whether to increase the number of resources in one resourceset or using multiple resource sets are candidate solution to achieve that, and do not need to be restricted for now.  3. For the “higher layer reporting”, as we are discussing the configuration of measurement resources, we do not need to mention about how to report.  For network-sided AI/ML model for BM-Case1 and BM-Case2, at least support using existing CSI framework to configure resources for Set A and Set B beams as the starting point.   * FFS on whether/how a measurement window or a number of measurement instances can be configured with the measurement resource set for BM-Case2. * FFS Whether/how to enable larger max number of measured resources for one CSI report ~~on whether to enlarge of the max number of resources in one~~ *~~CSI-ResourceConfig~~* * ~~Note: Higher layer based reporting is not precluded~~.   Note: the purpose, such as above “Set A and Set B beams” will not be specified in RAN 1 specifications ~~may or may not need to be explicitly defined/configured.~~ |
| Google | Support to reuse the CSI framework for the configuration |
| Apple | We also share a similar question as OPPO on set A.  As for set B configuration, we believe a key assumption is the Tx beams are hold unchangeable from one instance (one instance consists of 8 Tx beams for example, with Tx-1 to ~Tx-8) to another instance (again associated with Tx-1 to Tx-8). |
| TCL | Support to reuse the CSI framework for the configuration. The second FFS seems unnecessary as the measurement behaviors for AI purpose and the legacy are similar. |
| Ericsson | Generally fine. However, the proposal focuses on the configuration of set A/B, hence the following note is not clear “Note: Higher layer based reporting is not precluded”. We should focus on the configuration in this proposal, hence the note should be deleted. |
| LG | OK in principle, prefer to only have the main bullet and the last two sub-bullets regarding note. |
| QC | Further clarification is needed for first bullet: what is it exactly that cannot be achieved by reusing existing CSI framework? The second bullet (if needed) is also applicable to UE-side AI/ML models, so should not be discussed in this proposal. The proposal is about configuration of resources associated with Set A and Set B beams, so the first note (higher layer reporting) is not relevant to this proposal. With that said, we suggest the following update:  Updated proposal:  For network-sided AI/ML model for BM-Case1 and BM-Case2, ~~at least~~ support using existing CSI framework to configure resources for Set A and Set B beams as the starting point.   * ~~FFS on whether/how a measurement window or a number of measurement instances can be configured with the measurement resource set for BM-Case2.~~ * ~~FFS on whether to enlarge of the max number of resources in one~~ *~~CSI-ResourceConfig~~* * ~~Note: Higher layer based reporting is not precluded.~~ * FFS on whether further enhancements are needed beyond existing CSI framework.   Note: Association of resources to “Set A and Set B beams” may or may not need to be explicitly defined/configured. |
| CATT | Whether the proposal is applicable for training, inference, and monitoring? For inference and monitoring, the configurations of Set A may not be needed. For model inference of BM-Case2, the RS configuration can be enhanced when Set B=SetA (the details can refer to R1-2402366)  As UE may have different capability for RS measurement, the number of RS can be configured in one RS set may be less than the number of beams of SetA/SetB. The enhancement is needed for this case, e.g., configure multiple RS sets for Set A/Set B beams. |
| FL | My intention was for all purpose.  I don’t know whether there are company who want to explicitly configure “Set A” or “Set B” in the configuration? |
| ETRI | We generally fine with this proposal. |
| NEC | It may not be very clear on how to use the existing CSI framework to configure resources for Set A and Set B beams, we need to understand whether Set A and Set B beams are configured in different resource settings, different resource sets, different subsets, different groups or some other ways. |
| CMCC | Regarding to periodic or semi-persistent report, measurement window or the number of measurement instances does not need to be configured since gNB can decide measurement window by itself. Regarding to aperiodic report, configuration of measurement window can be considered.  Since beam number in set A may be larger than 64, and total RS for RSRP measurement across all *CSI-ReportConfig* may be larger than 64, extension of restriction on resource number in one *CSI-ResourceConfig* can be considered. |
| Panasonic | We are fine with the proposal. |
| InterDigital | We prefer to remove the FFS bullet. |
| Futurewei | We are fine with FL’s Proposal 2.1b |
| Lenovo | We are fine to use the existing CSI framework to support AI/ML model for BM-Case1 and BM-Case2. But we don’t think both Set A and Set B are needed to be configured, because the UE doesn’t need to know Set A and Set B for NW-side AI/ML model. |
| Xiaomi | Fine with the main bullet of the proposal 2.1b and also suggest to remove the FFS because it is related to reporting. and suggest to add the following FFS for some guidance.  FFS: whether to use one or two CSI-ReportConfig/ CSI-ResourceConfig/ resource set  FL: For NW sided model, I still cannot see the motivation to link two resources set in one CSI report |
| Ericsson | Also prefer to remove the FFS bullet. Xiaomi’s direction is ok if an FFS should be kept. |
| CATT | For the FFS, for model inference of BM-Case2, we think the RS enhancement can be considered when Set B=Set A.  The following figure shows an example of beam prediction of 2 future time instances based on 4 historic measurement results. When Set B=Set A, the RS is only transmitted in the measurement window, and the RS transmission overhead of prediction window can be saved.  In legacy, gNB can activate the semi-persistent RS transmission at the beginning of Measurement window1 and deactivate the RS transmission at the end of Measurement window1, and then activate the RS transmission at the beginning of Measurement window2 and deactivate the RS transmission at the end of Measurement window2, and so on. However, the overhead of activation/deactivation command will be large. We think the enhancement of RS activation can be considered. Moreover, this enhancement is needed for both NW-side model and UE-side model.  We suggest the following updates.  FFS on necessary enhancement, including whether/how a measurement window or a number of measurement instances can be configured with the measurement resource set for BM-Case2 at least for Set B=Set A. |
| CMCC | For second subbullet, it is not clear whether the intention is for configuration of setA or setA+setB? If the intention is to configure setA, we share similar view as HW, increasing maximum number of resources in one resourceset or using multiple periodic/semi-persistent resource sets under one CSI report can be considered. If the intention is to configure setA+setB, increasing maximum number of periodic/semi-persistent resourceset can be considered. |

## 2.2 For-UE sided model

|  |
| --- |
| 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 |

#### 2.2.1 Configuration of Set A

**Summary of proposals for Set A configuration**

* Alt 1: Set A reference signal,
  + Alt 1-1: Based on CSI framework
    - Ericsson [3] For UE-sided models, for the configuration of Set A, take the current CSI framework as the starting point
    - Vivo [6] use RS (SSB or CSI-RS) resource ID to configure Set A and/or Set B.
    - Samsung [13] where Set A is for prediction and Set B is a CSI-RS/SSB resource set for channel measurement.
    - InterDigital [11] RS resources for Set A including Set B are configured in a RS resource set.
    - Fujitsu [19] RAN1 to discuss the reference signal configuration of Set A, association between Set A and Set B, and possible reporting enhancement, taking the legacy CSI framework as starting point.
    - Nokia [25] Configure/indicate a second RS resource set associated with the CSI report configuration to consider as Set A. && Configure/Indicate a second RS resource set associated with the CSI report configuration.
    - ITL [26] 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
    - Fraunhofer HHI, Fraunhofer IIS [29] Set A of beams can be configured using the CSI reporting framework. FFS for further enhancements.
    - Qualcomm [34] a 2nd group of SSB/CSI-RS resources with prediction targets (i.e., Set-A beams), within a CSI report
  + Alt 1-2: Similar to the configuration of beams for BFR
    - Intel [5] consider explicit configuration of a set of resources for set A similar to the configuration of beams for BFR.
* Alt 2: Set A without reference signal
  + CATT [10] If the resource set of Set A is not configured to UE (e.g., for performance monitoring), the resource set of Set B and the AI model/functionality can be associated with the same CSI reporting.
  + Samsung [13] Set A consists of a set of predicted values defined by the size of Set A (K1) provided by the CSI-ReportConfig, e.g., Set A consists of the predicted integer values from 0 to K1 – 1.
  + Nokia [25]?? UE determines Set A based on QCL relations between Set B and configured CSI-RS resources.



CATT [10]

Figure 4 The association of Set B and Set A: (a) The resource set of Set A is configured; (b) The resource set of Set A is not configured

**Whether Set A needs to be measured or not**

* HW [1] Set A is not need to measure for inference
* OPPO [9] for UE sided model, support to configure Set A which may contain SSB resources and/or CSI-RS resource that UE doesn’t have to measure.
* Ericsson [3] 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

**FL0: Proposal 2.2.1b (Config for Set A for UE-sided model)**

For UE-sided AI/ML model for BM-Case1 and BM-Case2, further study for the following options considering the purpose for the configuration:

* Option 1: Using existing CSI framework to configure resources for Set A as a starting point
  + FFS necessary enhancement, including whether and when UE performs measurement for Set A.
* Option 2: Resources for Set A is not configured, e.g., only the size of Set A is provided.

**FL1: Proposal 2.2.1C (Config for Set A for UE-sided model)**

For UE-sided AI/ML model training, for BM-Case1 and BM-Case2, for the configuration of Set A, take current CSI framework as a starting point.

* FFS on whether to configure resources for Set A with more than one CSI-ResourceConfigId in one CSI-ReportConfig

For UE-sided AI/ML model monitoring, for BM-Case1 and BM-Case2, for the configuration of Set A, take current CSI framework as a starting point.

* FFS whether all resources for Set A needs to be configured
* FFS on whether to configure resources for Set A with more than one CSI-ResourceConfigId in one CSI-ReportConfig

For UE-sided AI/ML model inference, for BM-Case1 and BM-Case2, for the configuration of Set A, further study for the following options considering

* Option 1: Take current CSI framework as a starting point.
  + FFS necessary enhancement, including whether and when UE performs measurement for Set A.
  + FFS whether all resources for Set A needs to be configured
* Option 2: Resources for Set A is not configured, e.g., only the size of Set A is provided.

Note: Purpose, such as above “For UE-sided AI/ML model”, “training, monitoring, inference”, “for BM-Case1 and BM-Case2”, will not be specified in RAN 1 specifications

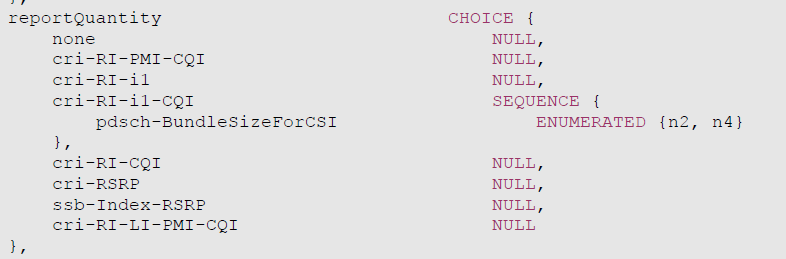
|  |  |
| --- | --- |
| Company | Proposals |
| FL | There are two directions I observed. I think this could be a good starting point for Set A configuration. The final downsselection may also related to Set B and Set A association, as well as consistency of additional condition. |
| NTT DOCOMO | It seems proponents of Option2 do not have the same solution in their mind, and each solution is proposed by only one company. Based on that, we suggest the following update.  For UE-sided AI/ML model for BM-Case1 and BM-Case2, further study:   * Option 1: Using existing CSI framework to configure resources for Set A   + FFS whether and when UE performs measurement for Set A. * ~~Option 2: Set A is configured without resources, e.g., only the size of Set A is provided.~~   Note: Other options are not precluded/ |
| New H3C | OK |
| HW/HISi | Option 1 During training, the UE needs to measure Set A. This should not be FFS for training. Also during training, since the UE needs to measure Set A, Option 2 should not be valid.  For inference, it may need to be studied further as suggested her.  Also, it could be included to study how to support large Set A that e.g. could not be included in one resource Set.  The proposal could be split into 2 proposal, 1 for training and 1 for inference:  **Updated Proposal 2.2.1a (Config for Set A for UE-sided model) for training**  For UE-sided AI/ML model for BM-Case1 and BM-Case2, further study for training:   * Option 1: Using existing CSI framework to configure resources for Set A   + ~~FFS whether and when UE performs measurement for Set A.~~   + FFS how to support large Set A that cannot be configured within one resource set * ~~Option 2: Set A is configured without resources, e.g., only the size of Set A is provided.~~   **Updated Proposal 2.2.1b (Config for Set A for UE-sided model) for inference**  For UE-sided AI/ML model for BM-Case1 and BM-Case2, further study at for inference:   * Option 1: Using existing CSI framework to configure resources for Set A   + FFS whether and when UE performs measurement for Set A.   + FFS how to support large Set A that cannot be configured within one resource set   Option 2: Set A is configured without resources, e.g., only the size of Set A is provided. |
| OPPO | Legacy configuration approach (SSB resources and/or CSI-RS resources) can be reused for Set A for UE-side model. The benefits of Option 1 when compared with Option 2 lies in the fact that UE may possibly measure the actual predicted SSB and/or CSI-RS resource(s) in Set A, hence UE may able to retune its Rx beam properly. But as for Option 2, UE only predicts index(es) without knowing its Rx beam(s).  With above being said, we support the change (i.e. removing Option 2) from DOCOMO. But we are not sure what would be other options. But fine to leave the open at early stage of Rel-19 WI. |
| Xiaomi | In our understanding, Option 1 can be used for model training and monitoring. While Option 2 can be used for model inference. We also suggest to discuss it for different purpose separately. i.e.,  Proposal 1: For UE-sided AI/ML model training and monitoring, using existing CSI framework to configure resources for Set A for BM-Case1 and BM-Case2.  Proposal 2: For UE-sided AI/ML model inference, for BM-Case1 and BM-Case2, further study:   * Option 1: Using existing CSI framework to configure resources for Set A   + ~~FFS whether and when UE performs measurement for Set A.~~ * Option 2: Set A is configured without resources, e.g., only the size of Set A is provided. * Option 3: Set A is not configured, e.g., only an ID is provided to indicate the Set A implicitly. |
| Fujitsu | Generally fine with the version from NTT DoCoMo.  One question just for clarification is whether the proposal could be applied for training data collection/inference/monitoring? |
| MediaTek | We prefer Option1. |
| ZTE | Prefer Option 1. The motivation to configure Set A with RS resources may at least include beam indication and performance monitoring. For performance monitoring, whether to transmit resources in Set A depends on the benchmark as will be discussed latter. For beam indication, however, the RS resources in Set A would be taken as the QCL source for beam indication, necessitating explicit prior configuration. |
| Google | We think we should clarify the report quantity. If this is for legacy L1-RSRP report, the NW can configure whatever RS. If this is for the report quantity based on the agreement in last meeting, e.g, beam information + predicted RSRP, we do not think set A should be configured. |
| vivo | We think from RS perspective, set A RS needs to be configured at least for training and monitoring. Further, for a certain beam report, a selection of set A resources needs to be indicated in the report configuration to configure the candidate beams for such report. |
| TCL | We support Opiton 1. It is preferred to define a clear mapping between Set A and Set B beams so the explicit indication of Set A is required. |
| Ericsson | Generally fine with the updated proposal from Xiaomi. Furthermore, it is not clear for us how only the size of set A could be used to configure set A during inference. Our view is that rather the NW would configure some “virtual” resources in the CSI framework that is not physically transmitted by the BW. We propose that we first agree on:  Proposal 1: For UE-sided AI/ML model training and monitoring, use existing CSI framework to configure resources for Set A for BM-Case1 and BM-Case2.  Proposal 2: For UE-sided AI/ML model inference, further study how to use existing CSI framework to configure resources for Set A for BM-Case1 and BM-Case2. |
| LG | It may be better to discuss training, inference, and monitoring aspects separately. It may also be better to merge with proposal 2.2.3(i.e., association between Set B and Set A). |
| QC | Do not agree with Option 2, as it is not clear how it works. Suggest the following update:  Updated Proposal 2.2.1  For UE-sided AI/ML model for BM-Case1 and BM-Case2, further study:   * Option 1: Using existing CSI framework to configure resources for Set A, as a starting point   + FFS whether additional enhancement is needed, including but not limited to, whether and when UE expects network transmission of Set A beams and when NW transmits Set A beams, whether and when UE performs measurement for Set A.   ~~Option 2: Set A is configured without resources, e.g., only the size of Set A is provided.~~ |
| CATT | We are ok with FL proposal. The motivation of proposal is to discuss how to configure/indicate Set A for UE-sided model. Option 1 can be used for model training or model inference, and option 2 can be used for model inference. We think both of option 1 and option 2 can be further study. |
| Ruijie | Support Option 1. |
| NEC | We think it is better to clarify whether Set A is configured or not in monitoring, inference and training, respectively? In principle, we think it is not needed to measure set A beams for inference.  Another comment is similar as our comment to Proposal 2.1 on the details of how to use existing CSI framework to configure set A. |
| CMCC | Option 1 is preferred. For both training and inference of UE-sided AI/ML model, configuration of Set A is necessary, the difference is whether Set A is measure or not. Thus, we prefer unify configuration of Set A across training and inference, option 2 is not suitable for data collection for training. |
| Panasonic | It is not clear how Option 2 works if only a size of set A is provided by NW. Option 2 introduces additional complexity for determining a measurement resource set and additional effort of measurements because UE may need to check many possibilities of the measurement resource set from all possible resources. On the other hand, Option 1 is simple and can work well. Therefore, we support Option 1. |
| InterDigital | We prefer Option 1 |
| Futurewei | We prefer Option 1. It is not clear to us how Option 2 can work for training. |
| Lenovo | We are fine with Option 1.  Option 2 is not clear to us. How can the UE obtain the beam information within a beam set by just having the size of the beam set? |
| Ericsson | Option 2 may be valid for the model inference scenario. However, given the general main text, option1 should be agreed as a starting point. Option 2 can be part of another proposal. |
| CATT | Option 2 can be considered for model inference phase.  If the performance monitoring is not based on the measurement results of Set A, the resource set of Set A will not be configured to UE. In this case, to represent the index of Set A beams, option 2 can be considered. For example, the set B can be associated with one AI model/functionality/configuration id, and then the UE will know the association and can report the index of predicted Top-K beams of Set A to gNB. |
| HW/HiSi 2 | Agree with E/// on preference for Option 1.  One further potential enhancement would be the support of a larger Set A, which could be captured in the FFS for enhancements of Option1 and potential discussion of Option 2.  Suggested update: FL0: **Proposal 2.2.1b (Config for Set A for UE-sided model)** For UE-sided AI/ML model for BM-Case1 and BM-Case2, further study for the following options considering the purpose for the configuration:   * Option 1: Using existing CSI framework to configure resources for Set A as a starting point   + FFS necessary enhancement, including e.g. whether and when UE performs measurement for Set A, whether and how to support a large Set A exceeding the size of one resource set.   ~~Option 2: Resources for Set A is not configured, e.g., only the size of Set A is provided.~~ |
| FL | As usual, no need to show preference for now. But pls ask question if some options are not clear |
| CMCC | For UE-sided AI/ML model training, there can be other options such as increasing maximum number of resources in one resourceset or using multiple periodic/semi-persistent resource sets under one CSI-ResourceConfig. |

##### FL2: **Proposal 2.2.1D (Config for Set A for UE-sided model)**

For UE-sided AI/ML model for BM-Case1 and BM-Case2, for the configuration of Set A, take current CSI framework as a starting point.

* For inference, Set A is implicitly or explicitly configured in the measurement report configuration for inference, where only resources for Set B needs to be measured.

|  |  |
| --- | --- |
| Company | Proposals |
| FL | Pls also indicate if you prefer other version for this issues |
| New H3C | OK in general |



##### FL2: **Proposal 2.2.1D (Config for Set A for UE-sided model)**

For UE-sided AI/ML model for BM-Case1 and BM-Case2, resources for Set A are configured, if applicable. FFS on the following options:

* Option A: use *CSI-ResourceConfig*
* Option B: introduce new resource set(s) outside of *CSI-ResourceConfig*
* Note: this does not precluded support both options.

For UE-sided AI/ML model for BM-Case1 and BM-Case2, support resources for Set B are configured, support

* Option A: use *CSI-ResourceConfig*
* ~~Option B: introduce new resource set(s) outside of~~ *~~CSI-ResourceConfig~~*
* Further study the following options when it needs to be used with potential downs election:
  + Opt 1: Link the *CSI-ResourceConfigId for Set A* to a *CSI-ReportConfig*
    - FFS on whether/when the measurements of Set A need to be reported
      * Note: when no measurements need to be reported, “reportQuantity” in the “CSI-ReportConfig” is set to “none”
  + Opt 2: Link the *CSI-ResourceConfigId for Set A* to a *CSI-ReportConfig* for inference
    - Where only resources for measurement needs to be measured
    - Note that, this does not configuration of *CSI-ResourceConfigId for Set A* in *CSI-ReportConfig* is not precluded.
  + Opt 3:
  + FFS on applicability of purpose for training, inference, monitoring

#### 2.2.2 Configuration of Set B

TBD

#### 2.2.3 Configuration of Set B and association with Set A

**FL0: Proposal 2.2.3b (Association between Set B and Set A)**

For association between Set A and Set B for UE sided model, at least when Set A and Set B are different or Set B is a subset of Set A, further study:

* Option 1: Implicitly,
  + Alt 1-1: via an indicator (data set ID, configuration ID, etc)
* Option 2: Explicitly
  + Alt 2-1: Resources for Set A is configured, and Set B is indicated as a subset of Set A
  + Alt 2-2: Resources for Set A and resources for Set B are configured in two separate Resources, configured, further study the following:
    - Alt A: With additional signalling to indicate the association
      * E.g., ID, mapping, indication if resource is part of set A/B,
    - Alt B: Without additional signalling to indicate the association
      * E.g., based on same NZP-CSI-RS-ResourceSetId, CSI-SSB-ResourceSetId across Set A and Set B
      * E.g., QCL information
  + Alt 2-3: Only resource Set B is configured,
    - with additional signalling to indicate the association
      * E.g., ID, mapping
* Combination of Option 1 and Option 2 are not precluded.
* Other options are not precluded.

**FL1: Proposal 2.2.3c (Association between Set B and Set A)**

For association between Set A and Set B for UE sided model, at least when Set A and Set B are different or Set B is a subset of Set A, further study:

* Option 1: Implicitly,
  + Alt 1-1: via an indicator (data set ID, configuration ID, etc)
* Option 2: Explicitly based on CSI framework,
  + Alt 2-1: In one CSI-ReportConfig, only resources for Set A is configured, and resources for Set B is indicated as a subset of resources for Set A,
    - E.g., bitmap
  + Alt 2-2: In one CSI-ReportConfig, both resources for Set A and resources for Set B are configured as two separate resources, further study the following:
    - Alt A: With additional signalling to indicate the association
      * E.g., ID, mapping, indication if resource is part of set A/B,
    - Alt B: Without additional signalling to indicate the association
      * E.g., based on same NZP-CSI-RS-ResourceSetId, CSI-SSB-ResourceSetId across Set A and Set B
      * E.g., based on same CSI-ResourceConfigId
      * E.g., QCL information
  + Alt 2-3: In one CSI-ReportConfig, only resource Set B is configured,
    - Alt A: with additional signalling to indicate the association
      * E.g., ID, mapping
      * Note: this assumes that resources for Set A are configured in a separate or resources for Set A is not configured.
    - Alt B: Without additional signalling to indicate the association
      * E.g., based on same NZP-CSI-RS-ResourceSetId, CSI-SSB-ResourceSetId across Set A and Set B
      * E.g., based on same CSI-ResourceConfigId
      * E.g., QCL information
      * Note: this assumes that resources for Set A are configured in a separate
* Combination of Option 1 and Option 2 are not precluded.
* Other options are not precluded.

|  |  |
| --- | --- |
| Company | Proposals |
| FL | Some companies mentioned to define a new QCL type. I think it may require configuration of Set A and Set B resources, and then link them via new QCL type. Not sure whether my understanding is correct or not.  If there are different alternatives, we can consider to expand the list. Pls provide your input.  BTW, example is not important, it can be removed if companies have concern. Currently, I hope the example can help you to better understand the proposal. |
| New H3C | OK |
| Spreadtrum | At what stage is the association configured, is it the training phase or the inference phase? Some companies seem to have different views on this. |
| HW/HiSi | Option 2 seems to be a precondition for Option 1. These options do not really seem to be orthogonal to each other. Can the “implicit method” please be further clarified with an example?  For example, when Set B is a subset of Set A, we need to define how beams from Set B can be mapped to Set A. This is a basic step included to be discussed firstly in our view. The data set ID mentioned in option 1 would then be a further step that could take place during a potential later discussion, e.g. when consistency is addressed.  Therefore, we think at this stage, we should only focus on how to map/associate Set A/B with each other.  For or understanding: can we discuss and clarify how a association without additional signalling would work?  Suggested update:  For association between Set A and Set B for UE sided model, at least when Set A and Set B are different or Set B is a subset of Set A, further study:   * ~~Option 1: Implicitly,~~    + ~~Alt 1-1: via an indicator (data set ID, configuration ID, etc)~~ * Option 2: Explicitly   + Alt 2-1: Resource for Set A is configured explicitly, and Set B is indicated as a subset of Set A   + Alt 2-2: Resource for Set A and resource for Set B are both configured,     - With additional signalling to indicate the association/mapping       * E.g., ID, mapping, association     - [Without additional signalling to indicate the association       * E.g., based on same NZP-CSI-RS-ResourceSetId, CSI-SSB-ResourceSetId across Set A and Set B       * E.g., QCL information]   + Alt 2-3: Only resource Set B is configured explicitly,     - with additional signalling to indicate the association       * E.g., ID, mapping, association * ~~Combination of Option 1 and Option 2 are not precluded.~~ * Other options are not precluded. |
| OPPO | We are in general okay to discuss the association between Set B and Set A. But it seems the association between Set B and Set A can be leveraged to address the consistency issue in Section 5. Hopefully, such dependence can be treated properly.  In addition, some minor comments we would like to share. First, Alt 2-1 seems not applicable for BM-Case2 where Set B can be the same as Set A, rather than a subset of Set A. For Alt 2-3, we think Set A should be anyway configured (discussed in previous proposals) to be associated with Set B. Hence, we suggest narrow it down before it becomes more complicated.   * Option 2: Explicitly   + Alt 2-1: Resource for Set A is configured explicitly, and Set B is indicated as a subset/full set of Set A   + Alt 2-2: Resource for Set A and resource for Set B are both configured,     - With additional signalling to indicate the association       * E.g., ID mapping     - Without additional signalling to indicate the association       * E.g., based on same NZP-CSI-RS-ResourceSetId, CSI-SSB-ResourceSetId across Set A and Set B       * E.g., QCL information   + ~~Alt 2-3: Only resource Set B is configured explicitly,~~      - ~~with additional signalling to indicate the association~~       * ~~E.g., ID mapping~~ |
| Xiaomi | As for implicit or explicit, we think Alt 1-1 is an explicit mechanism, but Alt 2-2 without addition signalling is an implicit mechanism.  In addition, as for Alt 2-2 with additional signalling, how does ID mapping indicate the association if Set B is different from Set A? if it can only be use in the case that Set B is a subset of Set A, why not use Alt 2-1?  Alt 2-3 can only be used for model inference? |
| Fujitsu | For the implicit option, could companies clarify how it works, for example, dataset ID?  Also share similar question with HW regarding “without additional signalling to indicate the association”. |
| MediaTek | We are generally OK, but we would like to know the background/motivation for Alt2-2: With additional signalling to indicate the association, why ID mapping is still required when resource for Set A and resource for Set B are already explicitly configured? |
| ZTE | Prefer Option 2 to indicate the association between Set A and Set B based on the existing CSI framework. Option 1 depends on the further progress in consistency assurance. |
| Google | We suggest adding an option to provide the configuration based on a beam ID from a beam codebook. |
| vivo | We think both options are needed, as they serve different purposes, or address different issues.   * Option 1 is actually used for alignment between training and inference, i.e., to make sure the inference uses an aligned data set with one certain training. * Option 2 includes alternatives on the configurations of set A and set B, including resource configuration and possible report configuration for set A. For example, what beams/resources in set A should be the candidates for a certain report.   Hence our thinking is it’s better and more clear to discuss these two issues separately. |
| TCL | We support option 2 to reuse the TCI state framework to link the Set A and Set B. |
| Ericsson | Similar **to 2.2.1,** we need to separate the training from the inference step. In inference, set A might be implicitly configured based on how the consistency is ensured. Hence, we first suggest focusing on the training step. It is further unclear what the example of “ID, mapping” comprises, our view is that the signaling should comprise an indication if the resource is part of set A or B: FL0: **Proposal 2.2.3 (Association between Set B and Set A)** For association between Set A and Set B for UE sided model training, at least when Set A and Set B are different or Set B is a subset of Set A, further study:   * ~~Option 1: Implicitly,~~    + ~~Alt 1-1: via an indicator (data set ID, configuration ID, etc)~~ * Option 2: Explicitly   + Alt 2-1: Resources for Set A is configured ~~explicitly~~, and Set B is indicated as a subset of Set A Resources   + Alt 2-2: Resources for Set A and resources for Set B are ~~both~~ configured in two separate Resources,     - With additional signalling to indicate the association       * E.g., ID, mapping, indication if resource is part of set A/B,     - [Without additional signalling to indicate the association       * E.g., based on same NZP-CSI-RS-ResourceSetId, CSI-SSB-ResourceSetId across Set A and Set B       * E.g., QCL information]   + ~~Alt 2-3: Only resource Set B is configured explicitly,~~      - ~~with additional signalling to indicate the association~~       * ~~E.g., ID mapping~~ * ~~Combination of Option 1 and Option 2 are not precluded.~~ * ~~Other options are not precluded.~~ |
| LG | We agree that this topic is important. However, it may not be clear which signalling option is implicit or explicit and some options seem only workable on special case (e.g., Alt2-1 is applicable only to the case where Set B is subset of Set A).  Before signalling details, another important aspect is ‘beam consistency in a resource’. Note that current specification allows NW to change the beam in different transmission instance even with a same resource ID without providing any information to UE, i.e. up to NW implementation. This means that even if Set A and Set B relation information is provided by NW per resource-level, NW can still change the beams per resource which will seriously degrade UE AI/ML beam predication accuracy. In this regard, Set A and Set B resources may be represented in a reference basis (e.g., 2D/3D coordinate) or beam IDs (which shall be consistent in any time). Or, some type of beam consistency related indication can be provided to UE. |
| QC | For Option 1, it is not clear how the association of Set B and Set A can be abstracted as an identifier. Identifiers such as dataset ID or configuration ID can be utilized to help with consistency of NW-side additional conditions across training and inference, but it is not quite clear how they can be used to describe the association between Set A and Set B beams.  Not clear what is meant by “additional signalling” in alt. 2-2. For instance, for QCL information, if a new QCL type is defined, there will be additional signalling associated with this new QCL type. Does this mean that new QCL information implicitly shows the association between Set B and Set A beams, and therefor there is no need for an explicit additional signalling? Also Alt. 2-3 is not clear and needs elaboration. |
| CATT | We are ok with FL proposal. The motivation of proposal is to discuss how to configure/indicate Set A for UE-sided model. Option 1 can be used for model training or model inference, and option 2 can be used for model inference. We think both of option 1 and option 2 can be further study. |
| NEC | 1. we also think using ID as in option 1 belongs to explicit method.  2. Among alternatives in Option 2, we prefer alt2-1 |
| CMCC | Option 1 is de-prioritized since it seems difficult for different gNB to align data set ID and configuration ID.  Alt2-3 is de-prioritized. As our comment to proposal 2.2.1, for both training and inference of UE-sided AI/ML model, configuration of (pre-configured) SetB and Set A is needed, the difference is whether Set A is measure or not. We prefer unify configuration of SetB and Set A across training and inference, Alt2-3 is not suitable for data collection for training.  Alt 2-1 is preferred with less signaling overhead, which can also be extended to the case Set A and Set B are different. For example, resourcesForChannelMeasurement (i.e. setB) can be a sequence of nzp-CSI-RS-ResourceSetList and csi-SSB-ResourceSetList. Nzp-CSI-RS-ResourceSetList corresponds to the case SetB is a subset, csi-SSB-ResourceSetList corresponds to the case Set A and Set B are different.  The following modification of Alt2-1 is suggested.   * + Alt 2-1: Resource for Set A is configured explicitly, and Set B is indicated as a different set or a subset of Set A     - E.g. Set B is indicated based on CSI-SSB-ResourceSetId or bitmap |
| Panasonic | We are supportive to discuss the association between Set B and Set A. For Option 1, can FL or proponent clarify how does it work? |
| InterDigital | We prefer Option 2. Among the alternatives, we prefer Alt 2-2. |
| Futurewei | It is not clear to us how Option 1 can work.  Regarding Option 2, we prefer Alt 2-1 and Alt 2-2. It is not clear to us how Alt 2-3 can work. |
| Lenovo | We want to add an example for option 1 as to associate a Set A for each of the Set B. Because there is no conclusion on the concept of ‘data set ID’, I suggest to remove the data set ID for option 1 at this stage. |
| Xiaomi | For Alt 1-1, one example is that two resource sets are configured, and the same configuration ID is configured to these two sets. That means one is set B (e.g., with lower set ID) and the other one is the associated set A.  For Alt B in Alt 2-2, does it mean UE can determine the set B and set A by QCL information?  For Alt 2-3, how does ID mapping work? |
| Ericsson | There seems to be a lot of confusion on what is meant by “association”.  In our view, it is simply the process of identifying which beams are part of set A/B. Propose that we first agree on the assumption on association/mapping. Then we can discuss methods on how to achieve it.  **Proposed conclusion: Association/mapping of Set A/B provides UE with information of which resources that are part SetA/B**  FL: How do we configure for the case when Set B is not part of Set A? |
| HW/HiSi 2 | Regarding the comment above from E///. In our view “mapping” is used when Set B is a Subset of Set A and indicates e.g. which beams from Set A are also Set B. Association, as there can be multiple Set A and multiple Set B, associates A set from Set A with a set from Set B.  We prefer option 2. For option 1 we agree with Lenovo for option 1 that we at least would like to see examples for data set ID and configuration ID. After this it could be discussed further.  In the off-line discussion, it would also be good if we can spend some time on to explain how the Alt 2-2/Alt B will work. This is not clear to us.  For Option 2   * Option 2: Explicitly   + Alt 2-1: Resources for Set A is configured, and Set B is indicated as a subset of Set A Resource   + Alt 2-2: Resources for Set A and resource for Set B are configured in two separate Resources, configured, further study the following:     - Alt A: With additional signalling to indicate the association       * E.g., ID, mapping, indication if resource is part of set A/B,     - [Alt B: Without additional signalling to indicate the association       * E.g., based on same NZP-CSI-RS-ResourceSetId, CSI-SSB-ResourceSetId across Set A and Set B       * E.g., QCL information]   + Alt 2-3: Only resource Set B is configured,     - with additional signalling to indicate the association       * E.g., ID, mapping |
| CMCC | Option 2 Alt2-1 and Alt2-2 are preferred, but the difference between option 2 Alt2-1 and Alt2-2A needs clarification. With Alt2-1, set B is indicated by bitmap, but resources for set B is configured as part of set A. |

##### FL2: **Proposal 2.2.3c (Association between Set B and Set A)**

For association between Set A and Set B for UE sided model, at least when Set A and Set B are different or Set B is a subset of Set A,

* CSI framework is used as a starting point, FFS on
  + Alt 2-1: In one CSI-ReportConfig, only resources for Set A is configured, and resources for Set B is indicated as a subset of resources for Set A,
    - E.g., bitmap
    - Note: Resources for Set A from multiple resource sets are not precluded.
  + Alt 2-2: In one CSI-ReportConfig, resources for Set A and resources for Set B are configured as separate resource sets, further study the following:
    - Alt A: With additional signalling to indicate the association
      * E.g., ID, mapping, indication if resource is part of set A/B,
    - Alt B: Without additional signalling to indicate the association
      * E.g., based on same NZP-CSI-RS-ResourceSetId, CSI-SSB-ResourceSetId across Set A and Set B
      * E.g., based on same CSI-ResourceConfigId
      * E.g., QCL information (TCI states)
  + Alt 2-3: In one CSI-ReportConfig, only resource Set B is configured,
    - Alt A: with additional signalling to indicate the association
      * E.g., ID, mapping
      * Note: this assumes that resources for Set A are configured in a separate CSI-ReportConfig or resources for Set A is not configured.
    - Alt B: Without additional signalling to indicate the association
      * E.g., based on same NZP-CSI-RS-ResourceSetId, CSI-SSB-ResourceSetId across Set A and Set B
      * E.g., based on same CSI-ResourceConfigId
      * E.g., QCL information (TCI states)
      * Note: this assumes that resources for Set A are configured in a separate CSI-ReportConfig
  + Other options are not precluded.

|  |  |
| --- | --- |
| Company | Proposals |
| FL | @CMCC, I think what your want is covered in Alt 2-2 Alt B |
| New H3C | OK in general |

# Report measurement and/or inference results

## 3.1 Report of inference results for UE sided model

|  |
| --- |
| **Agreement**  **For UE-sided model, at least for BM-Case1, for content in the report of inference results, support**   * **Opt 1: Beam information on predicted Top K beam(s) among a set of beams** * **Opt 2: Beam information on predicted Top K beam(s) among a set of beams and RSRP of predicted Top K beam(s) among a set of beams** * **At least K=1 and more, FFS on max value** * **FFS on beam information** * **FFS on the definition of predicted Top K beam(s)** * **FFS on definition of reported RSRP when applicable** * **FFS on other information in the report with potential down selection among the following options** * **Opt 3: Beam information on predicted Top K beam(s) among a set of beams and probability information of predicted Top K beam(s) among a set of beams**   + **FFS on the quantization method of probability information**   + **Probability information is the probability of the beam to be the Top 1 or Top K beam** * **Opt 4: Beam information on predicted Top K beam(s) among a set of beams, RSRP of predicted Top K beam(s) among a set of beams, and confidence information of the RSRP**   + **FFS on definition of reported RSRP**   + **FFS on the definition and quantization method of confidence information** * **Other options are not precluded.**   **where the set of beams is Set A, i.e., the beams for UE prediction.** |

#### Issue #1: Content of inference results

* Opt 3: Beam information on predicted Top K beam(s) among a set of beams and probability information of predicted Top K beam(s) among a set of beams
  + Yes:
    - Ericsson [3] 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
    - ZTE [7]
    - GOOGLE [8]: 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 [9] Probability information of predicted Top-K beam(s) could be quantized and reported to NW.
    - Samsung [13] the probability information could also be useful since the probability can reflect beam prediction accuracy in some extend
    - LGE [15] 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 [18] Support UE to report probability(ies) of predicted Top K beam(s) based on some pre-defined interval or threshold/criterion.
    - Sony [24] considering the content in the report of inference results, we support Options 1, 2, and 3.
    - Nokia [25] 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 [29] 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
  + FFS:
    - Huawei/HiSi [1] : if the probability information (Opt 3) of predicted Top-K beams would be reported, they should be quantized with overhead efficient manner.
* 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 [3] 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 [8]: 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.
    - LGE [15] 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 [18] 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.
    - Indian Institute of Tech (M), IIT Kanpur [37] When the AI/ML model output is L1-RSRP, the beam report should contain (1) RSRPs of the predicted Top *K* beam(s) among a set of beams, and (2) confidence information of the RSRPs (Opt 4 [1])
  + No:
    - Hw [1]: 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.

**FL: Hold on a little bit for performance monitoring**

#### Issue #2: FFS on beam information

Beam information for predicted beam:

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

* Ericsson [3], ZTE [7], Samsung [13]
* 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 from a configured codebook

* GOOGLE [8] Samsung [13] (for predicted beam)

**FL: to discuss this after agreement on Set A configuration.**

#### Issue #3: Definition of reported RSRP

For report content of inference results, the reported RSRP of predicted Top K beam(s) is:

* Predicted [L1-]RSRP, at least if the beam is not measured.
* FFS, if the beam is measured, considering:
  + Alt 1: Measured L1-RSRP;
    - vivo [6] ,ZTE [7], Xiaomi [16], Fujitsu [19]~~, DoCoMo [35] ??~~
  + Alt 2: predicted [L1-]RSRP
    - DOCOMO [35]
  + Alt 3: Both or one of them based on configuration
    - Alt 3-1: Measured L1-RSRP or predicted [L1-]RSRP based on configuration
      * CATT [10], CMCC [12]
    - Alt 3-2: Both or one of them with an indicator in the report.
      * Huawei/HiSi [1]
      * Intel [5] extra bit
      * Lenovo [22] 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.
      * Nokia [25] reported RSRP corresponding to CRI measured can be assumed to be measured RSRP, whereas reported RSRP corresponding to predicted CRI can be assumed to be predicted RSRP.

**FL0: Proposal 3.1.1b (definition of RSRP)**

For report content of inference results for UE-sided model, the reported RSRP of predicted Top 1 or Top K beam(s) is:

* Predicted RSRP**,** at least if the predicted Top 1 or Top K beam(s) is not measured.
* Further study the following alternative, if the predicted Top 1 or Top K beam(s) is measured,
  + Alt 1: Measured L1-RSRP;
    - Alt 2: Measured L1-RSRP and/or Predicted RSRP

**FL1: Proposal 3.1.1c (definition of RSRP)**

For report content of inference results for UE-sided model, for the reported RSRP of predicted Top 1 or Top K beam(s) further study the following options:

* Option A Predicted RSRP
* Option B: Predicted RSRP**,** if the beam is not measured, and measured L1-RSRP if the beam is measured
* Option C: both predicted RSRPand measured L1-RSRP if the beam is measured
* Combination of above options is not precluded.

|  |  |
| --- | --- |
| Company | Proposals |
| FL | I think we can define a new term as “predicted RSRP”.  If the predicted beam has been measured, further study two alternatives.  No company supports to report predicted RSRP only. |
| NTT DOCOMO | We do not support the proposal. If the predicted top K beam(s) is measured (in other words, predicted top-K beams include measured beams), the values of predicted RSRP is equal to the values of measured L1-RSRP with decent algorithm. In this case, it is not useful to differentiate measured RSRP or predicted RSRP, because the difference between measured RSRP and predicted RSRP is only the naming. We think proposal can be updated simply as follows.  For report content of inference results, the reported RSRP of predicted Top K beam(s) is:   * Predicted RSRP**~~,~~** ~~at least if the predicted Top K beam(s) is not measured.~~ * Note: If the predicted Top K beam(s) is measured, values of predicted RSRP should be equal to the values of measured RSRP. * ~~Further study the following alternative, if the predicted Top K beam(s) is measured,~~    + ~~Alt 1: Measured L1-RSRP;~~   + ~~Alt 2: Both measured L1-RSRP and Predicted RSRP, or one of them based on configuration~~     - ~~Alt 2-1: One of measured L1-RSRP or predicted RSRP based on configuration from NW~~     - ~~Alt 2-2: Both measured L1-RSRP and Predicted RSRP or one of them with an indicator in the report.~~ |
| New H3C | OK |
| Spreadtrum | It is worth considering that if there is an overall deviation between the predicted RSRP value and the real RSRP value, then the TopK beam can still be selected according to the predicted RSRP value. However, if the reported RSRP is mixed with the predicted value and the real value, it may interfere with the NW selection. |
| HW/HiSi | Seems fine. |
| OPPO | Assume Set B and Set A are both configured by NW, then NW can know whether the reported predicted beam has been measured (in Set B) or not (in Set A but not in Set B). Hence, for measured beam (in Set B), UE of course reports its measured L1-RSRP, whereas for predicted but not measured beam (in Set A but not in Set B), UE reports its predicted L1-RSRP.  Moreover, we cannot force the model to output the predicted L1-RSRP to be the same as measured L1-RSRP.  Then our suggestion is to go the easies way for L1-RSRP reporting.  For report content of inference results, the reported RSRP of predicted Top K beam(s) is:   * Predicted RSRP**,** at least if the predicted Top K beam(s) is not measured. * ~~Further study the following alternative,~~ if the predicted Top K beam(s) is measured,   + ~~Alt 1:~~ Measured L1-RSRP;   + ~~Alt 2: Both measured L1-RSRP and Predicted RSRP, or one of them based on configuration~~     - ~~Alt 2-1: One of measured L1-RSRP or predicted RSRP based on configuration from NW~~     - ~~Alt 2-2: Both measured L1-RSRP and Predicted RSRP or one of them with an indicator in the report.~~ |
| Xiaomi | We don’t see the motivation to report both of them for model inference. |
| Fujitsu | Since the inference results reporting is to report Top-K beams, the key aspect is how to determine the Top-K beams. Since there might be difference between the measured RSRP and the predicted RSRP, whether the measure RSRP or the predicted RSRP is used may impact on whether certain beam belongs to Top-K beams or not.  For the measured beams, we think whether predicted RSRP or measured RSRP is used could depend on the AI/ML model performance. If the AI/ML model performance is good, then the predicted RSRP could be used, otherwise the measured RSRP could be used.  We suggest the following update of the proposal.  *For report content of inference results of UE side model, the reported RSRP of predicted Top K beam(s) is when determining the Top-K beams:*   * *Predicted RSRP is used****,*** *~~at least if the predicted Top K beam(s)~~ if the beam is not measured.* * *Further study the following alternative, ~~if the predicted Top K beam(s)~~ if the beam is measured,*    + *Alt 1: Measured L1-RSRP is used;*   + *Alt 2: Predicted L1-RSRP is used;*   + *Alt 3: Measured L1-RSRP or predicted L1-RSRP is used depending on AI/ML model performance. If the performance is not good, the measured RSRP is used, otherwise the predicted RSRP is used.*   + *~~Alt 2: Both measured L1-RSRP and Predicted RSRP, or one of them based on configuration~~*     - *~~Alt 2-1: One of measured L1-RSRP or predicted RSRP based on configuration from NW~~*     - *~~Alt 2-2: Both measured L1-RSRP and Predicted RSRP or one of them with an indicator in the report.~~* |
| MediaTek | This proposal is only applicable when the RSRP is reported (i.e., Option 2 of the report content of UE side model inference), therefore, we suggest adding the following in the main bullet:  **For report content of inference results, when RSRP of predicted Top K beam(s) is reported, the reported RSRP ~~of predicted Top K beam(s)~~ is:** |
| ZTE | Prefer Alternative 1. When the beam to be reported originates from Set B, it is advisable to report the measured RSRP due to its higher reliability. The network can effortlessly discern between predicted and measured RSRP based solely on whether the reported beam belongs to Set B. On the contrary, for Alternative 2, reporting both the measured L1-RSRP and predicted RSRP would give rise to increased reporting overhead and variable reporting resource consumption, as the quantity of predicted Top K beams that have been measured remains unpredictable. |
| Google | Support the revision from Fujitsu |
| vivo | OK with the revision from Fujitsu. For the beams which are measured in set B, we don’t see any reason not to report the measured RSRPs, which are more reliable from gNB side. |
| KDDI | Support the revision from Fujitsu |
| TCL | We support to report both measured and predicted RSRP, especially when Set B is a subset of Set A beams, there will be a chance that the measured beam to be the optimal one. |
| Ericsson | We don’t see a strong need for this proposal. The predicted RSRP will probably anyway use the measured RSRP as input to improve the quality of the predictions. Hence our view is the following:  **Updated proposal for conclusion: When a measured RSRP is available, the NW can assume that at least such reported RSRP, possibly augmented with predictions, will fulfill existing measurement requirements from RAN4.** |
| LG | Agree with DOCOMO. In addition, we believe, in case that predicted top-K beam belongs to Set B, that predicted RSRP and measured RSRP shall be same or very similar so that it is not critical for NW to distinguish these. Therefore, we may further simplify the proposal as:  For report content of inference results, the reported RSRP of predicted Top K beam(s) is:   * Predicted RSRP**~~,~~** ~~at least if the predicted Top K beam(s) is not measured.~~ * ~~Further study the following alternative, if the predicted Top K beam(s) is measured,~~    + ~~Alt 1: Measured L1-RSRP;~~   + ~~Alt 2: Both measured L1-RSRP and Predicted RSRP, or one of them based on configuration~~     - ~~Alt 2-1: One of measured L1-RSRP or predicted RSRP based on configuration from NW~~     - ~~Alt 2-2: Both measured L1-RSRP and Predicted RSRP or one of them with an indicator in the report.~~ |
| QC | Agree with Docomo’s updated version. |
| CATT | We suggest to remove the detailed options for Alt2-2, there can be many detained options for Alt2-2. |
| NEC | For the proposal itself we are slightly prefer to report measured RSRP i.e., alt1.  In addition, it seems not very clear so that people may think this proposal means predicted RSRP only? No any resource indication (e.g. CSIRI). |
| CMCC | Ok. |
| Panasonic | We are fine with a version proposed by Fujitsu. |
| InterDigital | We can discuss this issue after having more fundamental designs. |
| Futurewei | We are in general fine with FL’s proposal with a slight preference on Alt 1. |
| Lenovo | Basically, we think the UE can report the predicted RSRP for the predicted top-K beams for the corresponding beam report when the reported beams are predicted beams by the AI/ML models. We are not sure whether additional procedure is needed for the UE to obtain the measurement RSRP for the predicted beam, because the NW need to send the predicted beams to the UE to obtain the measured RSRP after receiving UE’s feedback on the top-K predicted beams. |
| Xiaomi | Suggest to change Alt 2 to ‘predicted RSRP’ |
| Ericsson | Not supportive of Alt2, how would you test this alternative? |
| CATT | Suggest to change Alt 2 to ‘predicted RSRP’, and add Alt 3 as ‘predicted RSRP or measured RSRP based on gNB configuration’ |
| HW/HiSi 2 | Seems ok. |

##### FL2: **Proposal 3.1.1b (definition of RSRP)**

For report content of inference results for UE-sided model, the reported RSRP of predicted Top 1 or Top K beam(s) is:

* Predicted RSRP**,** at least if the predicted Top 1 or Top K beam(s) is not measured.
* Further study the following alternative, if the predicted Top 1 or Top K beam(s) is measured,
  + Alt 1: Measured L1-RSRP;
  + Alt 2: Measured L1-RSRP and/or Predicted RSRP

##### FL2: **Proposal 3.1.1c (definition of RSRP)**

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, further study the following options:

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

##### FL2: **Proposal 3.1.1D (definition of RSRP)**

(as working assumption)

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

##### FL2: **Proposal 3.1.1d (definition of RSRP)**

For report content of inference results for UE-sided model, for the reported RSRP of predicted Top 1 or Top K beam(s) is the predicted RSRP when the beam is not in Set B.

* Where the predicated RSRP is based on AI/ML output

|  |  |
| --- | --- |
| Company | Proposals |
| FL | Based on some offline discussion, I think in some implementation, the predicted RSRP is the same as measured L1-RSRP, or approximate to measured L1-RSRP.  Please think about it and indicate me which one is preferred. |
| New H3C | OK in general |

#### Issue #4: FFS on max value of K

TBD

#### Issue #5: Content of report for BM-Case 2

* Opt 1: one report of N(>=1) future time instances
  + HW [1]: model output compression, implicitly time stamp
  + Intel [5] For a UE-side AI/ML model, for BM-Case 2, the UE should report one or multiple predicted beams per time instance for the configured prediction window.
  + Vivo[6] support time stamp information in beam content for BM-Case2. time domain compression.
  + GOOGLE [8] , 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
  + OPPO [9] UE reports the predicted Tx beam for multiple future time instances by single reporting instance.
  + CATT [10] To reduce the reporting overhead, the temporal correlation of the predicted beams of multiple time instances can be considered.
  + CMCC [12] Regarding inference with UE-side model for BM-Case1 and BM-Case2, support L1 reporting of more than 4 beams and the associated L1-RSRP (if applicable) for at least one of N time instance(s) in one reporting instance.
  + Samsung [13] support CSI-ReportConfig to report predicted beams of N time instance(s) in one reporting instance
  + LGE [15] For temporal DL Tx beam prediction, information on time-variation of RSRP can also be included in the report.
  + NEC [18] study the method to configure the associated measurement and report resources for obtaining the ground truth in future time instances, the method to configure the associated measurement and report resources for obtaining the historical measurement results as model input.
  + ITL [26] it should be considered to report predicted beams of multiple future time instances in one reporting instance based on existing CSI-reportConfig as baseline
  + ETRI [27] for temporal domain beam prediction, support the CSI report format with the optimal K beam information along with RSRP information from multiple time instances.
  + DoCoMo [35] Since the probability of predicted top K beam(s) and the confidence of predicted RSRP represent the performance of beam prediction, the necessity of these information should be discussed in performance monitoring not in inference result reporting.
* Opt 2: Prediction results for one further time instance only.
  + NA

**FL0/FL1: Proposal 3.1.2b (Reports for BM-Case2)**

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

* FFS on the max total number of inference results over all instances

**FL0/FL1: Proposal 3.1.2c (Reports for BM-Case2)**

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.

|  |  |
| --- | --- |
| Company | Proposals |
| FL | Most of companies support the above proposals |
| NTT DOCOMO | Support the proposal. |
| New H3C | OK |
| Spreadtrum | OK |
| HW/HiSi | In principle, we do not have anything against this proposal. But before, it would be good to clarify about how many prediction results we are talking here.  Evaluations in the SI have shown that a large Top-K is needed for good performance and that Top-1 may not work. Assuming 4 prediction results per instance and N=4, would for example give 16 predictions results in the report. However, when discussing BM-case 1 we have not decided yet how large K should be chosen. If a large number of results is reported for BM-Case 2, shouldn’t this then also be feasible for BM-Case 1?  Suggested proposal:  For UE-side AI/ML model inference, for BM-Case2, ~~support~~ study the maximum number of ~~to report the~~ prediction results of N future time instance(s) to include in one reporting instance  FFS relationship to reporting Top-K beams for UE-side model in BM-Case 1. |
| OPPO | Support the FL proposal. |
| Xiaomi | Support the proposal and suggest to add one FFS:  FFS: support to report the prediction results of N future time instance(s) and the measured results of one history time instance in one reporting instance.  Since if Set B= Set A, the measured results of Set B can be reported to NW for reference. So the predicted results can be reported together with the measured results of the last history time instance.  FL: I pasted agreement wording. There is no intention to discuss the content in this proposal. |
| Fujitsu | Generally fine with the proposal.  Just some minor editorial update.  *For UE-side AI/ML model inference, for BM-Case2, support to report the prediction results of N (N>=1) future time instance(s) in one reporting instance* |
| ZTE | Support |
| Google | Support |
| Apple | Can the use case be clarified? e.g., under what situation such a feature is useful? |
| vivo | Support |
| KDDI | Support |
| TCL | Support |
| Ericsson | Share the view by HW that it is preferable to first put numbers on N time instances, and how many K beams to support. |
| LG | Support. |
| QC | OK |
| CATT | Support |
| Ruijie | Support |
| ETRI | We support this proposal. |
| NEC | Support |
| CMCC | Ok. |
| Panasonic | Support. |
| InterDigital | Support |
| Futurewei | Support. |
| Lenovo | Support |
| Xiaomi | What does ‘the content in the report of inference results’ mean? How about change to ‘the content of inference results’? |
| HW/HiSi 2 | In addition to the number of future instances, also the maximum number of total results to be included could be further studied  Suggested update:  For UE-side AI/ML model inference, for BM-Case2, support to report the content in the report of inference results of N(N>=1, FFS on N and on total number of inference results over all instances) future time instance(s) in one reporting instance |

## 3.2 Measurement report content 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 |

#### Issue #1: configuration and report of NW-sided model

##### FL2: Proposal 3.2A (Config and report for Set A and Set B)

For network-sided AI/ML model for BM-Case1 and BM-Case2, for the configuration and report of Set A and/or Set B, further study the necessary enhancements for the applicable purposes, including

* Enhancement for BM-Case2
* For Set A configuration,
  + Alt 1: whether to enlarge of the max number of resources per resource set (i.e., maxNrofNZP-CSI-RS-ResourcesPerSet)
  + Alt 2: whether to support more than one resource set in in one *CSI-ResourceConfig*
* whether to jointly or separately configure for Set A and Set B in one CSI-ReportConfig
* whether to jointly or separately configure two resource sets for Set A and Set B in CSI-ResourceConfig

|  |  |
| --- | --- |
| Company | Proposals |
| FL | Let’s have some discussion on necessary enhancement |
| New H3C | OK in general |

#### Issue #2: Report signalling

Summary from the contributions

* L1 based
  + Benefit:
    - Low latency, easily be reused HW/HiSi[1]
  + Huawei/HiSi[1], Spreadtrum [4]
  + Intel [5] for inference/monitoring
  + Vivo [6]
  + CATT [10] at least FFS in one or multiple reports
* MAC CE
  + Intel [5] for data collection
* RRC based
  + Spreadtrum [4]
  + Intel [5] for data collection
  + Vivo [6]
  + Samsung [13]
  + Xiaomi [16]
  + Fujitsu [19] Regarding training data collection for both UE side model and NW-side model, L3 signaling is preferred as the reporting container.
  + H3C [17]
  + ITL [26]
  + DoCoMo [35] Support data collection for NW side training performance monitoring via higher later signaling, where the detail mechanism should be up to RAN2.

**FL: Hold on for this meeting.**

## 3.3 Overhead reduction for measurement report

Summary of proposals of overhead reduction for a measurement report

* Method #1: Measurement omission/selection
  + Opt 1: Only report Top M beams with highest RSRP
    - HW/HiSI[1]
    - Ericsson [3] (NW sided model Set B) Support configuring reporting of at most N strongest set B beams.
    - ZTE [7] the maximum number of reported beam related information in one report can be configured by the NW based on UE capability indication.
    - CATT [10] Reporting Top-N L1-RSRP values;
    - CEWiT [30] support methods for reporting part of configured measured beams to reduce the size of report content for beam related information.
    - Sharp [36] ?? For NW-sided model training, support the number of beam measurement results including L1-RSRP and/or beam ID to be larger than 4 in a beam report. The number of beam measurement results is indicated by NW in its associated CSI report configuration.
    - Indian Institute of Tech (M), IIT Kanpur [37] UE reports beam-related information (for example L1-RSRP) corresponding to the top *N* beams
  + Opt 2: Only report the RSRP larger than a threshold
    - HW/HiSi[1]adaptive number of beams based on RSRP threshold.
    - Ericsson [3] ((NW sided model Set B) Support configuring reporting of only beams within X dB of the strongest beam,
    - ZTE [7]Data selection per sample (e.g., based on threshold criteria)
    - Xiaomi [16] support UE to indicate the RS IDs whose L1-RSRP are not reported because of lower than threshold to reduce overhead
    - NEC [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.
    - Apple [21] weak beams’ RSRPs can be omitted in the beam reporting for overhead reduction
    - ETRI [27] support the method of omitting RSRP values based on differences in RSRP values.
    - Fraunhofer HHI, Fraunhofer IIS [29] Adopt omitting of beams based on their strength.
    - CEWiT [30] support methods for reporting part of configured measured beams to reduce the size of report content for beam related information.
    - KDDI [32] methods to minimize the number of reporting beams as much as possible should be considered by using RSRP thresholds, etc.
  + Opt 3: Only report the RSRP with good channel condition
    - Ericsson [3] 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.
    - ZTE [7] Data omission among samples (e.g., according to data quality)
    - Interdigital [11] Support a sparse reporting mechanism for data collection for inference (e.g., based on channel conditions).
    - Fraunhofer HHI, Fraunhofer IIS [29] Study omission based on the measurement quality.
* Method #2: Reduce the overhead of beam information
  + Opt 1: bitmap-based
    - Huawei/HiSi [1] Whether the bitmap-based format is additionally adopted should be dependent on the range of M1 (i.e., number of reported beams) and M2 (size of Set A/Set B).
    - ZTE [7] 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.
    - Apple [21] 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.
  + Opt 2: No CRI/SSBRI ID
    - ZTE[7] 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.
    - OPPO [9] support to reduce the reporting overhead for both fixed or variable Set B, e.g. by dropping the part of SSBRIs/CRIs.
    - CATT [10] Alt.3: Omission the index of RS, e.g., reporting the L1-RSRP in order;
    - Interdigital [11] No CRIs/SSBRIs are reported and implicit beam indexes (e.g., by association with RSs and reported RSRPs) are used.
    - MTK [23] For the content of collected data for model inference of NW-side model, consider NW configures only Set B of beams for measurement and UE doesn’t report the resource indicators for the Set B of beams. B) The resources scheduled for measurement is the same as the resources to be reported, where the resource indicators can be discarded in the report.
    - NEC[18] including omitting CRI/SSBRI information.
    - Fraunhofer HHI, Fraunhofer IIS [29] depending on the configuration the UE can omit at least a part of the beam IDs to save reporting overhead.
    - CEWiT [30] 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.
  + Opt 3: based on pattern ID
    - Interdigital [11] *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.*
  + Opt 3: Reuse legacy CRI/SSBRI format
    - Huawei/HiSi [1]
    - Futurewei [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.
* Method #3: Quantization
  + Opt 1: Larger quantization step for reporting
    - Vivo [6] enhancements on quantization range and quantization step to reduce overhead for measurement results report.
    - CATT [10] Larger quantization step for reporting differential L1-RSRP.
    - MTK [23] quantizing L1-RSRP with lower number of bits than the current spec.
    - Nokia [25] consider enhancements for L1-RSRP quantization, increasing the differential L1-RSRPs in the report to X dB quantization step.
    - TCL [28] For overhead reduction purpose, study the quantization of report quantities, starting from the enhancement on the RSRP quantization.
    - Fraunhofer HHI, Fraunhofer IIS [29] Increased step sizes
    - CEWiT [30] support quantization of differential L1-RSRP with step size larger than 2dB to reduce the size of report content for beam related information.
    - KDDI [32] The number of representation bits and quantization step size of differential RSRP should be reconsidered.
    - DoCoMo [35] Large quantization step size for Set B measurement reporting
  + Opt 2: with differential L1-RSRP based reporting
    - Spreadtrum [4] existing quantitative criteria should be reused at least for model inference
    - Samsung [13] support differential L1-RSRP reporting for predicted beams.
    - Fraunhofer HHI, Fraunhofer IIS [29] Adaptive reference beam for differential RSRPs
  + Opt 3: reporting the normalized L1-RSRP measurement instead of actual L1-RSRP values (for NW sided model training?).
    - MTK [23]
* Method #4: At least for BMCase-2, report measurement results of multiple time instances in one report
  + Supported by
    - Huawei/HiSi [1]
    - Ericsson [3] support methods for UEs to compress the set B temporal domain measurement results to reduce the reporting overhead.
    - Vivo [6] time stamp information can be reported for BM-Case2. consider to use time domain data compression to reduce overhead.
    - OPPO [9]
    - CATT [10] report the measurement results of multiple time instances in one L1-RSRP reporting can be further considered for overhead reduction.
    - Interdigital [11] Reporting prediction results of multiple future time instances in one report should be supported.
    - Xiaomi [16] 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.
    - Apple [21] ?? 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.
    - MTK [23] For UE-side model inference, support reporting multiple time instances of beam prediction within one report. Further study on whether/how to explicitly and/or implicitly include corresponding time information in the report.
    - Nokia [25] , consider enhancements to report multiple past time instances in one reporting instance
    - DoCoMo [35] Consider the payload overhead reduction for UE reporting of predicted results at multiple time instances. Reporting of measurements from multiple time instances in one reporting instance.
  + Not supported by
    - **Spreadtrum [4]:** Including multiple past measurements in a single report can result in significant overhead but no improvement in AI performance. Conversely, if we report each time instance measurement, gNB can perform some analysis (e.g. model selection) based on the measurement results received first.

**FL0/FL1: Proposal 3.3 (overhead reduction)**

, further study the following methods with potential down selection for overhead reduction for the beam report in L1 signaling at least for content(s) for one time instance in one reporting:

* Method #1: Omission/selection of measurement quantity
  + Opt 1: Only report Top M beams with highest RSRP
  + Opt 2: Only report the RSRP larger than a threshold or within a threshold
    - FFS on how to define the threshold, e.g. an absolute value or a relative value compared to the highest RSRP
  + ~~Opt 3: Only report the RSRP with good channel condition~~
* Method #2: Reduce the overhead of beam information
  + Opt 1: bitmap-based
  + Opt 2: no CRI/SSBRI ID
  + Opt 3: based on pattern ID
  + Opt. 4: use a combinatorial indexing scheme to represent the selected strong beams
  + Note: the legacy CRI/SSBRI as a starting point
  + ~~Opt 4: Reuse legacy CRI/SSBRI format~~
* Method #3: Quantization of RSRP
  + Opt 1: Larger quantization step for L1-RSRP reporting
  + Opt 2: With differential L1-RSRP based reporting with a smaller quantization range
  + Opt 3: reporting the normalized L1-RSRP measurement instead of actual L1-RSRP values (for NW sided model training?).
  + Opt 4: Non-uniform quantization step size of RSRP.
  + Note: the legacy differential RSRP reporting as a starting point
* FFS on applicability for different purpose (e.g., training, inference, monitoring)
* FFS on applicability for NW-sided model and UE-sided model
* FFS on additional optimization for contents from multiple time instances in one reporting

|  |  |
| --- | --- |
| Company | Proposals |
| FL | In my understanding, Opt 2 of Method #1 can be a special case of Opt 3, any other example?  For BMCase-2, we can discuss it separately when there is clear agreement to support multiple time instances for some cases, e.g., prediction results from multiple time instances for UE-sided model. |
| New H3C | OK |
| Spreadtrum | The definition of beam information is not clear. |
| HW/HiSi | For method #1, another promising approach using a threshold would be to report all beams that are within a certain RSRP gap to the strongest beam.  For method #2, we think e.g. whether a bitmap is more efficient than legacy depends on the configuration to be used. Starting point should be legacy CRI/SSBRI for comparison.  For method #3, our understanding is that also legacy quantization for differential L1-RSRP reporting should be used as baseline for comparison.  Suggested update:  At least for NW-side model, further study the following methods for overhead reduction for the beam report in L1 signaling:   * Method #1: Omission/selection measurement quantity   + Opt 1: Only report Top M beams with highest RSRP   + Opt 2: Only report the RSRP larger than a threshold   + Opt 3: Only report the RSRP larger than a threshold relative to the highest RSRP   + ~~Opt 3: Only report the RSRP with good channel condition~~ * Method #2: Reduce the overhead of beam information   + Opt 1: bitmap-based   + Opt 2: No CRI/SSBRI ID   + Opt 3: based on pattern ID   Note: the legacy CRI/SSBRI as a starting point   * + ~~Opt 4: Reuse legacy CRI/SSBRI format~~ * Method #3: Quantization of RSRP   + Opt 1: Larger quantization step for L1-RSRP reporting   + Opt 2: With differential L1-RSRP based reporting   + Opt 3: reporting the normalized L1-RSRP measurement instead of actual L1-RSRP values (for NW sided model training?).   Note: the legacy differential RSRP reporting as a starting point   * FFS on applicability for different purpose (e.g., training, inference, monitoring) * FFS on applicability for UE-sided model |
| OPPO | We think it seems premature to include the mechanism of omission/selection. One good example is that for fixed Set B configuration, some resource(s) within Set B is by nature with low (lower than a threshold) L1-RSRP value, but it also provides useful information for NW-side model. Hence, we suggest to remove whole Method #1.   * ~~Method #1: Omission/selection measurement quantity~~   + ~~Opt 1: Only report Top M beams with highest RSRP~~   + ~~Opt 2: Only report the RSRP larger than a threshold~~   + ~~Opt 3: Only report the RSRP with good channel condition~~   We are open to discuss other Method(s). |
| Xiaomi | Support the proposal |
| Fujitsu | Clarification is needed. Is this proposal just for Set B measurement reporting for inference? |
| ZTE | For Opt 2 in Method #2, we suggest to modify it as ‘only CRI/SSBRI ID of the beam associated with the largest RSRP’. This revision is justified since, regardless of the scenario, the beam information associated with the largest RSRP should be reported as per the differential L1-RSRP based reporting mechanism.  Additionally, it appears that Opt 2 in Method #3 is not a parallel option with other options, as the differential L1-RSRP based reporting should be the foundation for all options.  The first FFS is not needed as we already agreed in the last meeting that the purpose of data collection will not be specified in RAN 1 specifications. |
| Google | For method #1, we suggest adding one option: It is up to UE implementation (same as legacy) |
| Apple | Support Huawei’s revision on Option 3 for method #1. And we suggest refining the wording as:   * + Opt 3: Only report the RSRP ~~larger than~~ within a threshold relative to the highest RSRP   For method #2: we can consider  Alt. 5: use a combinatorial indexing scheme to represent the selected strong beams. FFS details. |
| KDDI | Support |
| TCL | For Method #3, we suggest adding another option:  Opt 4: Non-uniform quantization step size of RSRP.  For example, the weak beams with low quality is considered as less significant, so it is quantized with larger step. |
| Ericsson | Support in general. For method 2, we should first clarify what is meant with beam information before listing options under method. Our suggestion is to keep it general before we have identified the content of such option.   * Method #2: Reduce the overhead of beam information   + ~~Opt 1: bitmap-based~~   + ~~Opt 2: No CRI/SSBRI ID~~   + ~~Opt 3: based on pattern ID~~   + ~~Opt 4: Reuse legacy CRI/SSBRI format~~ |
| LG | OK to further study but prefer to include “Opt 4: Reuse legacy CRI/SSBRI format” as a baseline for Method #2. |
| QC | From Spec perspective, we are considering overhead reduction methods for L1 signalling, so “at least for NW-side model” should be removed. Opt. 3 of Method#2 is not clear. Looks like this is related to variable Set B, but why does UE need to report the pattern ID, and why NW does not already know the associated pattern ID? Ned more clarification on the motivation for pattern ID. The proposal can be simplified as follows:  Updated Proposal 3.3:  ~~At least for NW-side model,~~ Further study the following methods for overhead reduction for the beam report in L1 signaling:   * Method #1: Omission/selection of measurement quantity   + Opt 1: Only report Top M beams with highest RSRP   + Opt 2: Only report the RSRP larger than a threshold   + ~~Opt 3: Only report the RSRP with good channel condition~~ * Method #2: Reduce the overhead of beam information   + Opt 1: bitmap-based   + Opt 2: No CRI/SSBRI ID   + Opt 3: based on pattern ID   + ~~Opt 4: Reuse legacy CRI/SSBRI format~~ * Method #3: Quantization of RSRP   + Opt 1: Larger quantization step for L1-RSRP reporting   + Opt 2: With differential L1-RSRP based reporting   + ~~Opt 3: reporting the normalized L1-RSRP measurement instead of actual L1-RSRP values (for NW sided model training?).~~ * FFS on applicability for different purpose (e.g., training, inference, monitoring) * FFS on applicability for NW-sided model and UE-sided model |
| CATT | OK |
| Ruijie | Support |
| Fraunhofer | Support in general. But we believe some clarification is required for Method#2 Opt 2. It has to be clear that Opt2 means in fact that the CRI/SSBRI can be transmitted for some beams, e.g. the strongest beam, but not for the remaining beams. |
| ETRI | As FL’s comment, this proposal discusses options for BM-case 1, one time instance in one reporting instance. Therefore, we suggest to update main bullet as follows.  At least for NW-side model, further study the following methods for overhead reduction for the beam report in L1 signaling for the case of one time in one reporting instance: |
| NEC | OK for this proposal and we think it is only for discussion of model inference. |
| CMCC | Ok. |
| Panasonic | Support. |
| vivo | Our first comment is this is applied for BM Case 1 as it restricts to report in one occasion. For BM Case 2, a different sets of candidate schemes can be listed for overhead reduction considering joint reporting of multiple occasions.  The second comment is the interpretation of “only” in some of these options. For example, in Method #1, to make Opt 2 work, UE needs to report a subset of RSRPs and its associated beam indices. Then this “only” is not accurate, as a report containing solely a subset of RSRPs does not work.  The last comment is to propose another option in Method #1:   * + Opt 3: Report beam/resource IDs with RSRP lower than a threshold   The motivation is if most of the beams in a measurement set is above this threshold, to report the beams lower than the threshold is more efficient in terms of overhead than the beams higher than the threshold. |
| Futurewei | Fine with the proposal. |
| Lenovo | OK |
| Ericsson | Not supportive of method 2. Can we first have a proposal on Issue #2? |
| HW/HiSi | Small updates suggestion for For Method #1:   * Method #1: Omission/selection of measurement quantity   + Opt 1: Only report Top M beams with highest RSRP   + Opt 2: Only report the RSRP larger than a threshold or within a threshold     - FFS on how to define the threshold, e.g. an absolute value or a relative value compared to the highest RSRP   + ~~Opt 3: Only report the RSRP with good channel condition~~ |

##### FL2 **Proposal 3.3 (overhead reduction)**

Further study the following methods with potential down selection for overhead reduction

Further discussed Omission/selection of measurement quantity for the beam report in L1 signaling at least for content(s) for one time instance in one reporting:

* Method #1: Omission/selection of measurement quantity
  + Opt 1: Only report Top M beams with highest RSRP
  + Opt 2: Only report the RSRP larger than a threshold or within a threshold
    - FFS on how to define the threshold, e.g. an absolute value or a relative value compared to the highest RSRP
* FFS on applicability of certain report ~~(e.g., a report for training, inference, monitoring)~~
* FFS on applicability for NW-sided model and UE-sided model
* FFS on additional optimization for contents from multiple time instances in one reporting
* Note: Method #2: Reduce the overhead of beam information and Method #3: Quantization of RSRP will be further discussed.

For overhead reduction, Method #1: omission/selection of measurement quantity for the beam report in L1 signaling can be considered at least for content(s) for one time instance in one reporting, including:

* Opt 1: Only report Top M beams with highest RSRP
* Opt 2: Only report the RSRP larger than a threshold or within a threshold
  + FFS on how to define the threshold, e.g. an absolute value or a relative value compared to the highest RSRP
* FFS on applicability of certain report ~~(e.g., a report for training, inference, monitoring)~~
* FFS on applicability for NW-sided model and UE-sided model
* FFS on additional optimization for contents from multiple time instances in one reporting
* Note: Method #2: Reduce the overhead of beam information and Method #3: Quantization of RSRP will be further discussed.
* Method #2: Reduce the overhead of beam information
  + Opt 1: bitmap-based
  + Opt 2: no CRI/SSBRI ID
  + Opt 3: based on pattern ID
  + Opt. 4: use a combinatorial indexing scheme to represent the selected strong beams
  + Opt 5: beam information omission according to probability information comparing to a threshold (only for inference for UE-sided model)
  + Note: the legacy CRI/SSBRI as a starting point
* Method #3: Quantization of RSRP
  + Opt 1: Larger quantization step for L1-RSRP reporting
  + Opt 2: With differential L1-RSRP based reporting with a smaller quantization range
  + Opt 3: reporting the normalized L1-RSRP measurement instead of actual L1-RSRP values (only for NW-sided model).
  + Opt 4: Non-uniform quantization step size of RSRP.
  + Note: the legacy differential RSRP reporting as a starting point
* FFS on applicability for different purpose (e.g., training, inference, monitoring)
* FFS on applicability for NW-sided model and UE-sided model
* FFS on additional optimization for contents from multiple time instances in one reporting

|  |  |
| --- | --- |
| Company | Proposals |
| FL | No change from FL1 |
| New H3C | OK in general |

##### FL2 **Proposal 3.3 (overhead reduction)**

Conclusion:

Further study the following methods for a reportQuantity for the beam report in L1 signaling at least for content(s) for one time instance in one reporting:

* For beam information
  + Opt 1: bitmap-based of configured resources for measurements
  + Opt 2: no CRI/SSBRI ID
  + Opt 3: based on pattern ID
  + Opt. 4: use a combinatorial indexing scheme to represent the selected strong beams
  + Opt 5: beam information omission according to probability information comparing to a threshold (only for inference for UE-sided model)
  + Note: the legacy CRI/SSBRI as a starting point
* Selection of quantity
  + Opt 0: related to All beams
  + Opt 1: Only report Top M beams with highest RSRP
  + Opt 2: Only report the RSRP larger than a threshold or within a threshold
    - FFS on how to define the threshold, e.g. an absolute value or a relative value compared to the highest RSRP

# 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 |

##### FL0/FL1/FL2: **Proposal 4 (Beam indication)**

For BM-Case2 (both UE-sided and NW-sided model), extend the Rel-17 TCI state activation/indication signalling 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 | Proposals |
|  | I didn’t list the supporting companies. But based on my reading, this got majority support. Other than the following companies  **Fraunhofer HHI, Fraunhofer IIS [29] Beam indication for multiple future time instances is not supported.** |
| NTT DOCOMO | Support the proposal. |
| Spreadtrun | Not support. |
| HW/HISi | Not support.  Our view was captured in our observation 6 “***Observation 6: For BM-Case 1 and BM-Case 2 with a NW-side model, the legacy TCI framework and mechanism are sufficient for handling beam indication.***”  We do not see the need for this proposal. In our view it is only applicable if Top-1 beams would be predicted. But SI evaluations have shown that Top-1 beam prediction is not sufficient, especially not for BM-Case 2. For Top-K prediction, we would have the normal P2 procedure.  Also, as mentioned, e.g. R1-2402366, if a scheduled PDSCH is sent, then there is no need to indicate the TCI state in a separate DCI. Thus, no overhead saving is achieved in case a PDSCH would be scheduled.  For the remaining cases (Top-1 prediction output, no scheduled PDSCH), the potential overhead saving seems insignificant. If our understanding is correct, then for e.g. for 80ms apart prediction instances and N=4, there would only be 3 DCIs saved within 320ms. With the proposal, one new DCI would be sent in the first prediction instance, compared to one DCI in each instance.  On the other hand, it will have a huge spec impact and also trigger subsequent studies (e.g. indicated TCI state overriding to allow some flexibility for the gNB). Furthermore, if we understand this proposal correctly, then this could mean that more TCI states need to be activated at the same time, which would be hard for UE implementation. If it on the other hand would mean to indicate non-activated TCI states, then the time lines need to be checked carefully for feasibility which would require involvement from RAN4. |
| OPPO | Support the FL proposal. |
| Xiaomi | Since only the DL Tx beam prediction is supported, UL Tx beam prediction is not supported. If separate TCI states are configured, it isn’t necessary to indicate N DL TCI state for N time instance but 1 UL TCI state for 1 time instance. So we suggest the following update:  For BM-Case2 (both UE-sided and NW-sided model), extend the Rel-17 TCI state activation/indication signalling methods to activate/indicate N joint TCI states which are corresponding to N future time ~~N~~ instances. |
| Fujitsu | Generally fine with the proposal. |
| MediaTek | We don’t support this proposal for the following reasons:   1. It is not mandatory and may require RAN4 involvement for the new time limitation for known/unknown TCI states. The standardization effort is huge while the potential benefit is not clear (see next point). 2. This proposal may potentially limit the capability of NW for choosing the Tx beam that can be shared by multiple UEs in each future N time instances, especially when there are legacy UEs coexist. Consider the following case where an AI/ML UE1 is static and another legacy UE2 is moving. If NW indicates future N TCI states at time 0 based on UE1’s prediction, NW will indicate beam 1 for all the future N time instances. However, since UE2 is moving, at time T(one of the future time instances), the beam that can simultaneously serve both UE1 and UE2 becomes beam2. In this case, NW still needs one DCI to activate beam2 at time T, then the additional DCI overhead for activating N TCI states at time 0 is wasted. |
| ZTE | Support |
| Google | Support |
| vivo | Support |
| TCL | Support |
| Ericsson | Similar views as above that this is probably not needed. First, we should identify the N number of time instances to be predicted, and the number of K beams to measure in each instance. Then we could estimate a possible overhead reduction by this proposal. The overhead reduction needs to be substantial in order to motivate the added complexity with the proposal. |
| LG | Support. |
| QC | Open to discussions in this direction, but first the need for such enhancement should be clarified for the temporal beam prediction use case. |
| CATT | Prefer not to extend the Rel-17 TCI state. Agree with HW that legacy TCI indication can be used for BM-Case 2. The necessity to extend legacy TCI state is not strong since the cases of DCI overhead saving is limited. For example, if there is PDSCH for transmission, the gNB will send DCI with DL assignment to UE, and the TCI field will always be present in DCI. In such a case, the DCI overhead cannot be reduced. |
| Ruijie | Support |
| Fraunhofer | Our concern is that any advantage provided by the indication of N TCI-States in the future is minimal for the signigicant complexity involved in modifying UE behaviour to apply the N TCI-States and the specification workload. So, we prefer to use the legacy TCI framework for TCI indication. |
| NEC | Support |
| CMCC | Ok |
| Panasonic | We suggest discussing this proposal later after concluding the proposals in earlier Sections (e.g., resource configurations, contents of measurements, etc.). |
| InterDigital | Support |
| Futurewei | The necessity of this proposal needs to be clarified first. |
| Lenovo | Not support at this stage because the benefit is not clear to us. |

# Consistency

## Consistency for UE sided model

|  |
| --- |
| TR 38.843 v2.0.0  ……  However, the AI/ML (without considering model switching) has significant performance degradation with some other unseen scenarios, including:  - For DL Tx beam prediction,  - deployment scenarios: UMi/Uma (at least with the assumption of different ISD, antenna height, down tilt and NLOS probability)  - various gNB setting: different gNB antenna array dimensions, and DL Tx beam codebook  - various Set B patterns  - various Set A patterns |

## Issue #1: Methods to ensure the consistency of additional condition for UE-sided model:

**FL0/FL1: Proposal 5.1 (Consistency)**

To ensure the consistency of NW-side additional condition across training and inference for UE-sided model, further study:

* Opt1: Based on an identifier
  + Opt1-1: global ID
  + Opt1-2: per-cell level identifier
  + FFS on the UE assumptions when receivingthe same identifier, including:
    - Same size of Set A of beams
    - Same size of Set B of beams
    - Same DL spatial TX-filter for each beam among all Set A of beams
    - Same DL spatial TX-filter ~~of~~ for each beam among all Set B of beams
    - Same indexing/ordering of all Set A of beams
    - Same indexing/ordering of all Set B of beams
    - Same Set B pattern(s) (i.e., association between Set A and Set B)
    - Same QCL assumptions for resources configured, including Set B and Set A
    - Same deployment scenarios at least for Opt1-1, e.g., ISD, antenna height, down tilt and NLOS probability
    - A model can be used, e.g., can meet the performance requirement.
    - Other options are not precluded.
  + FFS on how identifier is introduced, including:
    - Part of CSI framework
    - Other assumptions are not precluded
* Opt 2: Performance monitoring based, E.g., if a certain performance can be achieved.

|  |  |
| --- | --- |
| Company | Proposals |
| FL | First of all, model transfer is out of scope of this WI, and model identification is under study. In addition, opt1, the identifier could be a data set ID, model ID, configuration ID, etc. However, from my understanding, the name is not important. We can decide whether an identifier is needed or not. Then, whether this can be collected together with data for training (if specified), can be a second level of discussion.  It is encouraged for each company to provide some additional inputs on UE assumptions with the same identifier. Also, whether and how to specify the assumption. |
| NTT DOCOMO | We have concerns on UE and NW burden due to performance monitoring for consistency alignment.  Also, instead of using identifier, it is also possible to use consistency flag to represent whether consistency can be assumed or not without using identifier. In this way, the disclosure of proprietary information can be further relieved. This approach can be included as Opt3.  FL: What is the different from a flag to identifier, 1 bit vs N bits? |
| New H3C | OK |
| HW/HiSi | We think further discussion is needed firstly.  In our understanding, an identifier may not be needed. We can discuss the conditions/configurations firstly and then check if any additional conditions/identifiers are needed. For example, the consistency could be ensured by Set A/ Set B configuration and their mapping/association, no identifier would be needed for this. Based on this information, the UE-side can then autonomously determine which trained model to use.  Some comments regarding the FFS points in Option 1:   * Same size of Set A of beams   + Comment: This is not needed to be included in an identifier. The UE knows it already since it will be configured with Set A during training and inference * Same DL spatial TX-filter of all Set A of beams   + Comment: Could be ensured by NW implementation and UE model generalization capability. SI observation has shown that performance degradation is tolerable. * Same Set B pattern(s) (i.e., association between Set A and Set B)   + Comment: This is not needed to be included in an identifier. The UE knows it since it will be configured with SetB during training and inference and how they are mapped/associated * Same QCL assumptions for resources configured   + Comment: Should be visible from the configuration and is not needed as an identifier. * Same deployment scenarios at least for Opt1-1, e.g., ISD, antenna height, down tilt and NLOS probability   + Comment: This discloses proprietary and should be taken out or at least be handled very carefully. Also, we have not yet agreed if a global identifier (or an identifier at all is needed). It is very difficult for the network to know the granularity of these characteristics, which we should further discuss. * A model can be used, e.g., can meet the performance requirement.   + Comment: This should be up to the UE to decide, not for the NW identifier form the NW. Rather part of the monitoring outcome.   Suggested update:  To ensure the consistency ~~of additional condition~~ for UE-sided model, further study:   * Opt1: Based on an identifier   + ~~Opt1-1: global ID~~   + ~~Opt1-2: per-cell level identifier~~   + ~~FFS~~ on the UE assumption ~~with the same identifier~~, including:     - Same size of Set A of beams     - Same DL spatial TX-filter of all Set A of beams     - Same Set B pattern(s) (i.e., association between Set A and Set B)     - Same QCL assumptions for resources configured     - FFS if/how to describe Same deployment scenarios at least for Opt1-1, e.g., ISD, antenna height, down tilt and NLOS probability, without disclosing proprietary information     - ~~A model can be used, e.g., can meet the performance requirement.~~     - Other options are not precluded.   Note: Proprietary information needs   * Opt 2: Performance monitoring based, E.g., if a certain performance can be achieved |
| OPPO | We are in general fine to address the consistency issue via an identifier. The detail UE assumption with such identifier can be further studied. The consistency can be ensured by configuring such an ID before the inference phase. But as for Opt 2, the monitoring-based approach can only be applicable when the AI/ML model enters the inference phase. It is late and bring up additional latency due to the monitoring and decision-making procedure.  Hence, if possible, we suggest to remove Opt 2.   * ~~Opt 2: Performance monitoring based, E.g., if a certain performance can be achieved.~~ |
| Xiaomi | With per-cell level ID, does it mean UE need to support different model for different cell even the addition condition is same between different cells?  Regarding Opt 2, we share same concern as DOCOMO on the UE burden. |
| Fujitsu | We don’t agree to remove Opt 2.  The purpose to ensure consistency is to guarantee the performance. Hence, the performance monitoring based is feasible way. |
| ZTE | For the listed UE assumptions, some of them can be enabled by the current spec without specifying any further identifier, such as the size of Set A of beams and the QCL assumptions for resources configured. Besides, it would be quite challenging for specifying global ID without disclosing proprietary information, due to the need for coordination across gNB or infrastructure vendors, for example, for the alignment of DL spatial TX-filter of all Set A of beams or deployment scenarios.  Considering the NW-side additional condition is still conceptual, RAN1 should first discusses concrete NW-side additional conditions specific to AI/ML beam management and then discuss the potential consistency approaches if necessary. |
| Google | OK |
| Apple | Support the FL proposal to start discussion on this critical issue. For Option-2, we don’t think that is feasible: due to the small coverage of FR2, performance monitoring based approach won’t be fast enough to be of use to AI/ML BM’s inference operation. |
| vivo | Prefer FL’s formulation. It is needed to list the detailed options. |
| TCL | We support Opt-1. |
| Ericsson | We agree with the view that an identifier is needed to ensure consistency. The meaning of such identifier needs to be well-studied to ensure a good identifier definition. Where the identifier is valid could be part of another discussion, however, it is important to discuss how it can be introduced. Our suggestion is the following clarification update: **Updated Proposal 5.1 (Consistency)** To ensure the consistency from training to inference of NW-side additional condition for UE-sided model, further study:   * Opt1: Introducing an identifier   + ~~Opt1-1: global ID~~   + ~~Opt1-2: per-cell level identifier~~   + ~~FFS on the~~ UE assumptions when receiving the same identifier, including:     - Same size of Set A of beams     - Same DL spatial TX-filter of all Set A of beams     - Same Set B pattern(s) (i.e., association between Set A and Set B)     - Same QCL assumptions for resources configured     - Same deployment scenarios at least for Opt1-1, e.g., ISD, antenna height, down tilt and NLOS probability     - A model can be used, e.g., can meet the performance requirement.     - Other options are not precluded.   + How identifier is introduced, including:     - Part of CSI framework     - Other assumptions are not precluded * Opt 2: Performance monitoring based, E.g., if a certain performance can be achieved. |
| LG | Agree that this issue needs to be resolved and prefer Opt1. However, there are several questions/comments:   * Opt1-2: why this needs to be ‘per-cell’? This could be revised to ‘local ID’ to be competitive to Opt1-1 but we feel that it is not critical whether it is global or local at this stage. * On FFS   + Unclear on the meaning/necessity of ‘Same QCL assumptions for resources configured’   + Unclear on the meaning/necessity of ‘A model can be used, e.g., can meet the performance requirement.’ |
| QC | When we say “same” in the sub-bullets, we need to be clear what it means, that’s why “across training and inference” is added in the main proposal text.  Updated Proposal 5.1  To ensure the consistency of NW-side additional condition across training and inference for UE-sided model, further study:   * Opt1: Based on an identifier   + Opt1-1: global ID   + Opt1-2: per-cell level identifier   + FFS on the UE assumption with the same identifier, including:     - Same size of Set A of beams     - Same size of Set B of beams     - Same DL spatial TX-filter ~~of~~ for each beam among all Set A of beams     - Same DL spatial TX-filter ~~of~~ for each beam among all Set B of beams     - Same indexing/ordering of all Set A of beams     - Same indexing/ordering of all Set B of beams     - Same Set B pattern(s) (i.e., association between Set A and Set B)     - Same QCL assumptions for resources configured, including Set B and Set A     - ~~Same deployment scenarios at least for Opt1-1, e.g., ISD, antenna height, down tilt and NLOS probability~~     - ~~A model can be used, e.g., can meet the performance requirement~~.     - Other options are not precluded.     - Note: Each beam for Set A is associated with a corresponding resource index during training and inference, respectively. Given same identifier and same resource index, the same Set A beam should be used during training and inference. The same principle holds for Set B. * ~~Opt 2: Performance monitoring based, E.g., if a certain performance can be achieved.~~ |
| CATT | Do not support Opt1.  The per-cell level identifier will not solve the consistency issue for inter-cell scenario. For intra-cell scenario, the identifier is not necessary, since we can assess whether the UE-sided model is matched to the “scenario/configuration” via performance monitoring.  For inter-cell case, how to assign the “identifier” and whether this is workable considering inter-cell/region/vendor/PLMN cooperation are not clear. We think the approach may not be feasible in practice |
| NEC | Consistency based on monitoring i.e., option 2 is quite complicated and we don’t think it is the right direction to pursue. |
| CMCC | The size of set A, size of set B, association between Set A and Set B can be aligned through configuration of set A and set B. Regarding to other aspects such as DL spatial TX-filter of all Set A of beams, it is difficult to align global ID among all gNB. With per-cell level identifier, when UE is in mobility, how can UE know two different identifiers in two cells correspond to the same gNB setting? |
| Panasonic | We think that to achieve consistency between training and inference for UE-sided model, it is required to discuss to conclude two following questions:   * Q1: What kind of NW-additional condition is needed to achieve consistency between training and inference? * Q2: How to ensure consistency between training and inference?   Proposal 5.1 is referred to Q2, but Q1 is not concluded. To save effort, we think we should conclude Q1 first before discussing Q2. |
| Futurewei | We share the same view as other companies that it would be better to firstly discuss what kind of NW-side additional condition is needed to ensure consistency across training and inference before directly jumping into the signaling details. |
| FL | Discuss what additional condition is one way. Another way is, how do we want to handle additional information. And how it can work. Via ID, what kind of information can be and need to be resolved.  I think it is quite clear based on the study item outcome. |

##### FL2: **Proposal 5.1 (Consistency)**

To ensure the consistency of NW-side additional condition across training and inference for UE-sided model, further study:

* Opt1: Based on an identifier
  + Opt1-1: global ID
  + Opt1-2: per-cell level identifier
  + FFS on the UE assumptions when receiving the same identifier, including:
    - Same size of Set A of beams
    - Same size of Set B of beams
    - Same DL spatial TX-filter for each beam among all Set A of beams
    - Same DL spatial TX-filter ~~of~~ for each beam among all Set B of beams
    - Same indexing/ordering of all Set A of beams
    - Same indexing/ordering of all Set B of beams
    - Same Set B pattern(s) (i.e., association between Set A and Set B)
    - Same QCL assumptions for resources configured, including Set B and Set A
    - Same deployment scenarios at least for Opt1-1, e.g., ISD, antenna height, down tilt and NLOS probability
    - A model can be used, e.g., can meet the performance requirement.
    - Other options are not precluded.
  + FFS on how identifier is introduced, including:
    - Part of CSI framework
    - Other assumptions are not precluded
* Opt 2: Performance monitoring based, E.g., if a certain performance can be achieved.

|  |  |
| --- | --- |
| Company | Proposals |
| FL | No updates |
| New H3C | OK in general |

## 5.2 Consistency for NW sided model

## Issue #1: Whether UE Rx assumption needs to be specify or not?

Summary from the contributions:

For UE Rx assumption for NW-sided model:

* Alt 1: No additional specification enhancement
  + Spreadtrum [4]: no additional spec enhancements are required to ensure the common understanding between NW and UE on Rx beam assumption.
* Alt 2: Some specification is needed for Rx beam information
  + Ericsson [3], Fixed Rx-beam during data collection Indicate Rxbeam index during data collection
  + Intel [5] 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 [6] 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. ; To maintain consistency of QCL assumption across training and inference phases, UE may assume the same rule for QCL assumption, e.g. apply the best/fixed/same Rx beam(s).
  + CATT [10] ?? it is beneficial to align the Rx information of the measurements between network and UE.
  + Xiaomi [16] 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.
  + Fujitsu [19] 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).
  + DOCOMO [35] If NW side beam prediction gets difficult due to UE side additional condition (e.g., UE Rx beam assumption), some enhancements should be introduced.

##### FL0/FL1/FL2: **Question 5.2 (UE Rx for NW-sided model)**

Further study on UE Rx assumption for NW-sided model for consistency during data collection and inference

* Alt 1: No additional specification enhancement
* Alt 2: Additional specification is needed

|  |  |
| --- | --- |
| Company | Proposals |
| FL | Q1: What is current Rx beam assumption in existing CSI report? and whether it can be reused?  Q2: If cannot be reused, or if some enhancement is needed, what is the issue if using existing Rx beam assumption?  Q3: What is proposed solution to handle UE Rx assumption? For example,   * A) using a fixed Rx beam of all the measurements in a report * B) indicating Rx beam information in data collection (what kind of information? I don’t think index makes sense since NW doesn’t know the meaning of index) * C) others? |
| New H3C | OK |
| OPPO | We would prefer Alt.1, since Rx beam is up to UE implementation and Rx beam prediction has been deprioritized due to such reason in SI phase.  Q1: Take the SSB-based beam sweeping in current CSI report, it is reasonable for UE to assume the best Rx beam for reported SSB as Tx beam. In our view, it can be reused.  Q2: Other than fixed Rx beam. There could be other options, e.g. best Rx beam or up to UE implementation. Hope it helps to include more options.   * C) using best Rx beam of all the measurements in a report * D) the Rx beam(s) is (are) up to UE implementation |
| Xiaomi | For the beam report in existing CSI report, current Rx beam assumption is up to UE implementation. But for beam prediction, different assumption during measurement report for training and inference may result in low prediction accuracy. |
| Fujitsu | It’s impossible to use the same UE Rx beam between training data collection and inference since training data collection and inference are different stage.  But for training data collection, we think the same Rx beam should be applied for the measurement on Set B and corresponding Set A.  Also for performance monitoring, we think the same Rx beam should be applied for the measurement on Set B and corresponding Set A. |
| ZTE | The current Rx beam assumption for measurement is indicated by the QCL relations. Specifically, at least for the fixed Rx beam assumption of all the measurements in a report, it can be achieved by configuring all resources in the resource set with the same QCL type D relation, which indicates implicitly that the same spatial Rx beam should be used for measurement. Open to discuss the alignment of other Rx beam assumptions. |
| Google | Support Alt1 |
| vivo | We support Alt 2. In the SI, there are simulation results showing if Rx beam is just based on one Tx beam in P3 measurement following current spec, there will be performance loss. Hence we think the current spec needs some enhancement so that the Rx beam determination in P3 measurement can have more Tx beams to measure. |
| TCL | We support Alt 1. |
| Ericsson | Support to study |
| QC | Support Alt. 1 |
| CATT | Alt 2  Q1：in current CSI report，the Rx beam is up to UE implementation  Q2：if reuse existing Rx beam assumption, UE may use different Rx assumption during measurement for training and inference; the low prediction accuracy will be low.  Q3: c) based on reporting by UE |
| NEC | We support further study the spec impact if needed, i.e., alt 2. |
| CMCC | Fine to reuse current Rx beam assumption. That is, gNB configures Rx beam based on prior information, which can be seen as quasi-optimal Rx beam.  If best Rx beam is pursued, some enhancement may be needed to let UE perform complete Rx beam sweeping instead of using the same Rx beam as QCL source. |
| Panasonic | Support to further study. |
| InterDigital | We support Alt. 2. |
| Futurewei | We prefer Alt 1. |
| Lenovo | We think the current QCL framework is sufficient at least for inference operation. |

# Performance monitoring (TBD)

## Issue #1: Metric and benchmark

Summary from the contribution.

* Alt.1: Beam prediction accuracy related KPIs, e.g., Top-K/1 beam prediction accuracy
  + **Yes:** Huawei/HiSi [1] Intel [5], vivo [6], ZTE [7] as baseline, OPPO [9], CATT [10], Samsung [13], xiaomi [16], H3C [17], Fujitsu [19], Lenovo [22], ITL [26], [32]
* Alt.2: Link quality related KPIs, e.g., throughput, L1-RSRP, L1-SINR, hypothetical BLER
  + **Yes:** Huawei/HiSi [1] Intel [5], ZTE[7], Nokia ??[25], Futurewei
  + **Precluded**: Spectrum [4] ( it is difficult to determine the performance predicted by AI based on this metric alone)
* Alt.3: Performance metric based on input/output data distribution of AI/ML
  + **Deprioritized/precluded**: Huawei/HiSi[1] Spectrum [4] ( Alt.3 has not been studied enough to be used as an effective metric to monitor performance.)
  + **Yes:** 
    - Ericsson [3] metric 3 is needed if UE does not provide any confidence/uncertainty of its predictions.
    - Samsung [13]
    - NEC [18] Monitoring based on data distribution should be supported
    - DoCoMo [35]?? Since the probability of predicted top K beam(s) and the confidence of predicted RSRP represent the performance of beam prediction, the necessity of these information should be discussed in performance monitoring not in inference result reporting.
* Alt.4: The L1-RSRP difference evaluated by comparing measured RSRP and predicted RSRP
  + **Deprioritized/precluded**: Huawei/HiSi[1]
  + **Yes:** Intel [5],Vivo [6], CATT [10] (baseline), Ericsson [3](if reported predicted L1-RSRP) , Samsung [13], Xiaomi [14] ,H3C [17], Fujitsu [19] H3C [17], Lenovo [22], Sony [24], ITL [26] KT [31], KDDI [32]

**FL: Although Alt 1 and Alt 4 have majority support from companies, I still have some questions on how to use it. The next step is to mapping the metric to performance monitoring procedure.**

##### FL2: **Proposal 6.1 (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
* Alt.1-2: Hypothetical on beam prediction accuracy related KPIs, e.g., Top-K/1 beam prediction accuracy, based on configured resource(s)
* 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 3-1: input data distribution, any definition??]~~
* Alt 4-1: The L1-RSRP difference between the measured [L1-]RSRP and predicted RSRP according to beam(s) in the same target Set A resources, 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)
* Alt 4-2: The L1-RSRP difference between measured [L1-]RSRP of current beam and predicted RSRP of the predicted Top 1 beam

##### FL2: **Proposal 6.2 (monitoring):**

**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: 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)**
* **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:**

* **All above alternatives**

**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 model selection/activation/ deactivation/switching/fallback operation:**

* **All above alternatives**

**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**

|  |  |
| --- | --- |
| Company | Proposals |
| FL | Q1: any other metric to be added or clarified?  Q2: any additional mapping the metric(s) and monitoring options? |

## Issue #2: Performance monitoring procedure

|  |
| --- |
| **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  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. |

Proposal:

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

- 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)

* 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, e.g., Top-K/1 beam prediction accuracy, based on configured resource(s)
* 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 3-1: input data distribution, any definition??]
* Alt 4: The L1-RSRP difference between the measured [L1]-RSRP and predicted RSRP

- 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

Metrics:

* 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, e.g., Top-K/1 beam prediction accuracy, based on configured resource(s)
* 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 3-1: input data distribution, any definition??]
* Alt 4: The L1-RSRP difference between the measured [L1]-RSRP and predicted RSRP

# Others

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

* HW: legacy mechanism of CSI processing unit and CSI processing timeline can be reused as a starting point
* Vivo [6] further study whether to define AI process capability including re-use or modified the existing CSI computation time and CSI processing units.
* Lenovo [22] Consider to introduce AI process units for AI based operation.
* MTK [23] For UE-sided model, consider how to adapt current beamReportTiming framework/definition to include the AI/ML’s model inference delay.
* [??] For UE-sided models, for inference, study the UE reporting its inference time to the gNB.

## Issue #2: Whether/how to address Measurement error

* Ericsson [3] The number of samples and statistical metrics of the performance metrics needs to be addressed.
* Intel [5] 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 [8] 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 [19] 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.

**FL: Low priority. Only handle this if all other issues have good progress.**

# 8 Proposals for online

##### **Proposal AAA(Consistency)**

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 at least impacts UE assumption on beams of Set A/Set B:

* Opt1: Based on associated ID (in AI 9.1.3.3)
  + FFS on what can be assumed by UE with the same associated ID
  + FFS on how associated ID is introduced, e.g., within CSI framework, or outside of CSI framework
  + FFS on how NW assigns the associated IDs
* Opt 2: Performance monitoring based, E.g., whether a certain performance can be achieved.
* Other options are not precluded.
* Note: this does not precluded support both options, or combined one.
* FFS on further details on other necessary NW-side additional condition
  + Note: the above options can be applicable for other necessary NW-side additional condition (if identified),
* FFS on whether the above options can be applicable for other conditions
* ~~Note: “NW-side additional condition at least includes: gNB setting (e.g., different gNB antenna array dimensions, and DL Tx beam codebook, indexing/ordering of Set A/Set B beams)”~~

##### **Proposal ZZZ (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
* 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 BBB (NW-sided model, beam information in the report):**

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

* Opt 0: legacy CRI/SSBRI
* Opt 1: beam indexes are reported based on a bitmap, where bitmap indicating RS index of a resource set. L1-RSRPs are reported for indicated bitmap.
* Opt 2: No beam index reporting. 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. 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.
* Opt 5: Index of a group of beams (identified as subset resource set of a resource set) is reported with all L1-RSRPs of the group of beams.
* Other options are not precluded.

# References

1. R1-2402023 Discussion on AI/ML for beam management Huawei, HiSilicon
2. R1-2402054 Discussion on specification support for AI/ML-based beam management FUTUREWEI
3. R1-2402056 AI/ML for beam management Ericsson
4. R1-2402094 Discussion on AIML for beam management Spreadtrum Communications
5. R1-2402144 Specification support for AI/ML for beam management Intel Corporation
6. R1-2402230 Specification support for beam management vivo
7. R1-2402263 Discussion on specification support for AI/ML beam management ZTE
8. R1-2402276 AI/ML based Beam Management Google
9. R1-2402316 On specification for AI/ML-based beam management OPPO
10. R1-2402366 Specification support for AI/ML-based beam management CATT
11. R1-2402609 Discussion on AI/ML for beam management InterDigital, Inc.
12. R1-2402553 Discussion on specification support for beam management CMCC
13. R1-2402491 Discussion for supporting AI/ML based beam management Samsung
14. R1-2402626 Discussion on specification support for beam management Panasonic
15. R1-2402628 Discussions on AI/ML for beam management LG Electronics
16. R1-2402649 Specification support for beam management Xiaomi
17. R1-2402730 Discussions on AIML for beam management New H3C Technologies Co., Ltd.
18. R1-2402756 Discussion on specification support for beam management NEC
19. R1-2402786 Discussion on specification support on AI/ML for beam management Fujitsu
20. R1-2402846 Specification support for AI-enabled beam management NVIDIA
21. R1-2402869 On AI/ML for beam management Apple
22. R1-2402918 AI/ML specification support for beam management Lenovo
23. R1-2402939 Discussion on specification support for AI/ML-based beam management MediaTek
24. R1-2402957 Discussions on AI/ML for beam management Sony
25. R1-2402996 AI/ML for Beam Management Nokia
26. R1-2403006 Specification support for AI/ML beam management ITL
27. R1-2403011 Discussion on specification support for beam management ETRI
28. R1-2403036 Discussion on AI/ML beam management TCL
29. R1-2403041 Specification support for beam management Fraunhofer HHI, Fraunhofer IIS
30. R1-2403051 Discussion on Specification Support for Beam Management CEWiT
31. R1-2403131 Discussion on AI/ML based beam management KT Corp.
32. R1-2403141 Specification support for beam management KDDI Corporation
33. R1-2403157 Discussions on AI/ML for beam management CAICT
34. R1-2403182 Specification support for AI-ML-based beam management Qualcomm Incorporated
35. R1-2403232 Discussion on AI/ML for beam management NTT DOCOMO, INC.
36. R1-2403299 Discussions on specification support for beam management Sharp
37. R1-2403367 Discussions on Specification Support of AI/ML for Beam Management Indian Institute of Tech (M), IIT Kanpur

# 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

*Tue AM*

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