**3GPP TSG RAN WG1 Meeting #110e-bis R1-2210xxx**

**eMeeting, October 10 – 19, 2022**

**Source:** Moderator (Samsung)

**Title:** Feature lead summary #1 evaluation of AI/ML for beam management

**Agenda Item:**  9.2.3.1

**Document for:** Discussion and Decision

# Introduction

In RAN#94-e, Rel-18 new study item on “Study on Artificial Intelligence (AI)/Machine Learning (ML) for NR Air Interface” was approved. The following use cases were identified as the initial set:

* Initial set of use cases includes:
  + CSI feedback enhancement, e.g., overhead reduction, improved accuracy, prediction [RAN1]
  + Beam management, e.g., beam prediction in time, and/or spatial domain for overhead and latency reduction, beam selection accuracy improvement [RAN1]
  + Positioning accuracy enhancements for different scenarios including, e.g., those with heavy NLOS conditions [RAN1]

The performance of AI/ML based algorithms for the use cases includes the following aspects:

1. Evaluate performance benefits of AI/ML based algorithms for the agreed use cases in the final representative set:
   * Methodology based on statistical models (from TR 38.901 and TR 38.857 [positioning]), for link and system level simulations.
     + Extensions of 3GPP evaluation methodology for better suitability to AI/ML based techniques should be considered as needed.
     + Whether field data are optionally needed to further assess the performance and robustness in real-world environments should be discussed as part of the study.
     + Need for common assumptions in dataset construction for training, validation and test for the selected use cases.
     + Consider adequate model training strategy, collaboration levels and associated implications
     + Consider agreed-upon base AI model(s) for calibration
     + AI model description and training methodology used for evaluation should be reported for information and cross-checking purposes
   * KPIs: Determine the common KPIs and corresponding requirements for the AI/ML operations. Determine the use-case specific KPIs and benchmarks of the selected use-cases.
     + Performance, inference latency and computational complexity of AI/ML based algorithms should be compared to that of a state-of-the-art baseline
     + Overhead, power consumption (including computational), memory storage, and hardware requirements (including for given processing delays) associated with enabling respective AI/ML scheme, as well as generalization capability should be considered.

In this contribution summarized the discussions and proposal on evaluation methodology (EVM) and KPIs from contributions submitted to AI 9.2.3.1 for beam management (BM). The issues that are in the focus of this round of the discussion are furthermore tagged FL2.

Follow the naming convention in this example:

* Document-v000-Mod.docx
* Document-v001-Mod-CompanyA.docx
* Document-v002-CompanyA-CompanyB.docx

If needed, you may “lock” a spreadsheet file for 30 minutes by creating a checkout file, as in this example:

* CompanyC uploads an empty file named Document-v003-CompanyB-CompanyC.checkout
* CompanyC checks that no one else has created a checkout file simultaneously, and if there is a collision, CompanyC tries to coordinate with the company who made the other checkout
* CompanyC then has 30 minutes to upload Document*-v003-CompanyB-CompanyC.docx*
* If no update is uploaded in 30 minutes, other companies can ignore the checkout file.

To avoid excessive email load on the RAN1 email reflector, please note that there is NO need to send an info email to the reflector just to inform that you have uploaded a new version of this document. Companies are invited to enter the contact info in the table below.

#### FL1: Question 0-1

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

|  |  |  |
| --- | --- | --- |
| Company | Point of contact | Email address |
| Ericsson | Henrik Ryden | Henrik.a.ryden@ericsson.com |
| vivo | Hao Wu | hao.wu@vivo.com |
| Samsung | Jackson Wang  Yeon-Geun Lim | h0809.wang@samsung.com  yg.lim@samsung.com |
| CATT | Da Wang | wangda@catt.cn |
| Huwei, HiSilicpn | Thorsten Schier | Thorsten.schier@huawei.com |
| NVIDIA | Xingqin Lin | xingqinl@nvidia.com |
| Qualcomm | Hamed Pezeshki | hamedp@qti.qualcomm.com |
| OPPO | Jianfei (Jeffrey) Cao | caojianfei@oppo.com |
| Lenovo | Srinivas Kothapalli | vkothapalli@lenovo.com |
| MediaTek | Gyubum Kyung Yu-Jen Ku | gyubum.kyung@mediatek.com  yu-jen.ku@mediatek.com |
| Xiaomi | Mingju LI | limingju@xiaomi.com |
| ZTE | Wenfeng Liu | liu.wenfeng@zte.com.cn |
| CMCC | Jun Zuo | zuojun@chinamobile.com |
|  |  |  |

# General evaluation assumptions

## 1.1 (closed)Open issues on evaluation assumption of SLS

The following proposals were discussed in contributions:

**gNB antenna configuration and transmission power**

Proposals in contributions:

* ZTE [3]
  + Proposal 1: Unified descriptions of the antenna configuration for BS and UE should be adopted to avoid confusion.
    - BS antenna configuration: antenna setup and port layouts at gNB: [4, 8, 2, 1, 1, 1, 1], (dV, dH) = (0.5, 0.5) λ
    - UE antenna configuration: antenna setup and port layouts at UE: [1, 4, 2, 1, 2, 1, 1], 2 panels (left, right)
* Google [9]
  + Proposal 1: For EVM, the BS antenna configuration should be (M, N, P, Mg, Ng) = (4, 8, 2, 1, 1), (dV, dH) = (0.5, 0.5) λ.
  + Proposal 2: Add BS height = 10m as a second option as evaluation assumption to be aligned with evaluation assumption in other agenda items and to create more beams for indoor UEs in vertical domain.
* Samsung [24]
  + Proposal # 1: Adopt the following parameter for BM SLS evaluation:
    - BS Antenna Configuration and BS Tx Power
      * Antenna setup and port layouts at gNB: [4, 8, 2, 1, 1,1,1], (dV, dH) = (0.5, 0.5) λ as baseline with 40dBm Tx power
* Qualcomm [26]
  + Proposal 1: Consider the following simulation assumptions for BM-Case1 and BM-Case2:
    - BS antenna configuration: [8, 16, 2, 1, 1,1,1], (dV, dH) = (0.5, 0.5) λ
    - BS Tx power: 34 dBm

#### FL1: (closed)Antenna configuration and DL Tx power

**Proposal 1-1-1a:**

* **BS antenna configuration:** 
  + **antenna setup and port layouts at gNB: [4, 8, 2, 1, 1, 1, 1], (dV, dH) = (0.5, 0.5) λ**
  + **Other assumptions are not precluded**
* **BS Tx power:** 
  + **40dB or 34 dBm reported by companies**
  + **Other values are not precluded**
* **UE antenna configuration (Clarification of agreement in RAN 1 #110):** 
  + **antenna setup and port layouts at UE: [1, 4, 2, 1, 2, 1, 1], 2 panels (left, right)**
  + **Other assumptions are not precluded**

|  |  |
| --- | --- |
| Supporting companies | Futurewei, Google, MediaTek, LG, Xiaomi, OPPO,NTT DOCOMO, Lenovo, vivo, Samsung, ZTE, NVIDIA |
| Objecting companies |  |

**Please provide your view Proposal 1-1-1a, if any.**

|  |  |
| --- | --- |
| Company | Comments |
| Google | We think BS height = 10m should be added to be aligned with other WI, 25m is a bit high so that there is less beam with the vertical direction smaller than 90 degree. |
| HW/HiSi | For the BS TX, what is the motivation to add another value to the agreed BS TX power of 40 dBm?  If we in general should strive for converged number across companies (as it is the case for other parameters), we could take one value as baseline and then add another one as optional:   * **BS Tx power:**    + **40dBm ~~or 34 dBm reported by companies~~**   + **Other values (e.g. 34 dBm) are not precluded and can be reported by companies** |
| Nokia | BS Tx power should be set to not exceed the EIRP, given the number of antenna elements in the array. For instance, if companies agree that EIRP should not exceed 73 dBm (Table A.2.1-1 in 38.802), then BS Tx power 40 dBm is fine with 4x8 ant array (EIRP = 63 dBm) but also with 8x16 array (EIRP = 69 dBm). Not sure why multiple values of BS Tx power are required to list here. |
| Lenovo | It would be good to agree on one number for the Tx power. |
| Ericsson | We would like to avoid having an excessive number of simulation alternatives. We share the view of HW/HiSi and their updated proposal. |
| Samsung | We prefer 40dBm of BS Tx power as baseline. |
| CATT | Share the same view with HW/HiSi and their updated proposal. |
| FL1 | Based on the current discussion, the proposal was updated for GTW on Monday.  **Proposal 1-1-1b:** (updated)   * **BS antenna configuration:**    + **antenna setup and port layouts at gNB: (4, 8, 2, 1, 1, 1, 1), (dV, dH) = (0.5, 0.5) λ**   + **Other assumptions are not precluded** * **BS Tx power:**    + **40dBm**   + **Other values (e.g. 34 dBm) are not precluded and can be reported by companies** * **UE antenna configuration (Clarification of agreement in RAN 1 #110):**    + **antenna setup and port layouts at UE: (1, 4, 2, 1, 2, 1, 1), 2 panels (left, right)**   + **Other assumptions are not precluded** |
| FL1 | **Agreement**   * **BS antenna configuration:**    + **antenna setup and port layouts at gNB: (4, 8, 2, 1, 1, 1, 1), (dV, dH) = (0.5, 0.5) λ**   + **Other assumptions are not precluded** * **BS Tx power for evaluation:**    + **40dBm (baseline)**   + **Other values (e.g. 34 dBm) are not precluded and can be reported by companies** * **UE antenna configuration (Clarification of agreement in RAN 1 #110):**    + **antenna setup and port layouts at UE: (1, 4, 2, 1, 2, 1, 1), 2 panels (left, right)**   + **Other assumptions are not precluded** |

**Traffic model**

**Agreement in RAN 1 #109e**

|  |  |
| --- | --- |
| **Traffic Model** | FFS:   * Option 1: Full buffer * Option 2: FTP model   Other options are not precluded |

Proposals in contributions:

* Huawei/HiSi [2]
  + Proposal 8: For the selection of the traffic model for beam prediction, **full buffer** considered as the starting point.
* Interdigital [6]
  + Proposal 12: For traffic model, support the following evaluation assumptions:
    - For beam information related KPIs, no traffic model is needed to be defined as UE is only measuring reference signals not decoding actual PDSCHs.
    - For system performance related KPIs, FTP traffic should be used to reflect practical traffics for the evaluation.
    - For FTP traffic model, FTP model 3 is preferred as generating a new UE for each packet (FTP model 1) is not appropriate for evaluating benefits from AI/ML based beam prediction.
  + Proposal 13: For UE distribution, support the following evaluation assumptions:
    - For FTP traffic model, 10 UEs per cell/sector with 50% and 70% RUs is preferred.
    - 80% outdoor UEs and 20% indoor UEs for spatial domain beam prediction as defined in TR 38.901 (Option 1).
* LGE [10]
  + Proposal 1. FTP model 1 with packet size of 0.5 Mbytes can be considered as a baseline traffic model.
  + Proposal 2. If FTP model 1 is selected for the baseline traffic model, consider RU of 30%, 50%, 70%, and companies are required to report the assumption of load factor for each of RU values.
* Intel [14]
  + Proposal 4: For SLS UE distribution, large number of UEs per cell should be allowed for dataset generation but should be limited to 10 UEs/TRP for throughput evaluation using trained model for beam selection.
  + Proposal 5: For system performance KPIs, if supported, only full-buffer traffic models should be used.
* Samsung [24]
  + Proposal # 1: Adopt the following parameter for BM SLS evaluation:
    - Traffic Model
      * Option 1: Full buffer
      * Other options are not precluded

#### FL1: (closed)Traffic model

**Proposal 1-1-2a:**

* **For system performance related KPI (if supported) [e.g, throughput] evaluation (model inference), companies report the traffic model:**
  + **Option 1: Full buffer**
  + **Option 2: FTP model with detail assumptions (e.g., FTP model 1, FTP model 3)**

|  |  |
| --- | --- |
| Supporting companies | Futurewei, Google, MediaTek, OPPO, NTT DOCOMO, Lenovo, Ericsson, vivo, NVIDIA |
| Objecting companies |  |

**Please provide your view Proposal 1-1-2a, if any.**

|  |  |
| --- | --- |
| Company | Comments |
| LG | Is this proposal intended to agree both options as baseline? |
| Xiaomi | Same question as LG, and suggest following update  **Proposal 1-1-2a:**   * **For system performance related KPI (if supported) [e.g, throughput] evaluation (model inference), companies report one of the following traffic model:**   + **Option 1: Full buffer**   + **Option 2: FTP model with detail assumptions (e.g., FTP model 1, FTP model 3)** |
| Nokia | Support |
| CATT | Prefer to down select one. We think full buffer can be as a starting point. |
| FL1 | **Conclusion**   * **For system performance related KPI (if supported) evaluation (model inference), companies report either of the following traffic model:**   + **Option 1: Full buffer**   + **Option 2: FTP model with detail assumptions (e.g., FTP model 1, FTP model 3)** |

## 1.2 (Closed) Trajectory model for UE mobility

Companies provide views on the three options for UE trajectories:

* Huawei/HiSi [2]:
  + Proposal 11: For the evaluation of temporal domain beam prediction, Option 4, i.e., random direction straight-line trajectories for randomly dropped UEs, should be considered as the starting point.
* ZTE [3]:
  + Proposal 5: The random direction straight-line trajectories in Option #4 can be adopted for modelling UE trajectory, which is simpler than other UE trajectory options and beneficial for model generalization.
* Intel [14]:
  + UE trajectories with straight line movement without sharp turns should be considered as a first step for evaluation.
  + The UE trajectory should be sampled at least at the minimum decorrelation distance of the large-scale parameters corresponding to the scenario of evaluation.
* Nokia [19]:
  + Proposal 12: RAN1 further investigates the trajectory model for BM-Case#2, adopting Option #4 as a starting point for further studies.

FL0: There is no strong need to down select the baseline performance in this meeting other than current agreements.

## 1.3 Others

Other than the open issues for SLS, the following proposals were proposed by companies:

* Vivo [5]:
  + It is encouraged for companies to provide publicly accessible dataset and disclose the details for the dataset generation as much as possible for training and validation for cross-check purposes.
* FL0: We already have sufficient agreements for dataset. There is no need to have further agreements, especially under the discussion of sub-use case.
* Ericsson [11]:
  + Observation 6 For beam prediction evaluations consider providing the results with measurement accuracy noise modelled as additive gaussian noise with 95% of the density function within the measurement accuracy range, and/or uniformly distributed noise
  + Proposal 7 Study the impact of measurement imperfections on model performance for the considered beam prediction use cases.
  + Proposal 8 Consider the following to mitigate the L1-RSRP measurement inaccuracy impact in ML based beam prediction
    - Possibility to tighten requirements on L1-RSRP measurement accuracy
    - Define different UE capability based on their capability in fulfilling a measurement accuracy requirement.
* Mediatek [20]:
  + Observation 1: Both machine learning models perform better on ray-tracing dataset compared to SLS dataset.
  + Proposal 2: Study and evaluate the performance of AI/ML beam prediction using the dataset generated by the ray-tracing simulations.
* NVIDIA [23]
  + Proposal 1: Companies are encouraged to contribute real data to develop and evaluate AI/ML based algorithms for beam management.
* FL0: The above proposals can be covered by the agreements in framework. No need for further discussion.
* Samsung [24]:
  + **Data collection:**
    - 8 RBs as baseline, companies can report larger number of RBs
    - First 2 OFDM symbols for PDCCH, and following 12 OFDM symbols for data channel
  + **Channel model:**
    - LOS channel: CDL-D/E extension,
    - NLOS channel: CDL-A/B/C extension,
    - CDL-D extension, DS = 100ns as baseline.
    - Companies explains details of extension methodology considering spatial consistency.
    - Other channel models and DSs are not precluded.
* Qualcomm [26]:
  + Proposal 2: For both sub use cases BM-Case1 and BM-Case2, clarify interpretation of “set B” by selection of one of the following alternatives
    - Alt.1: Set B is a set of beams, whose measurements are performed (for prediction of set A)
    - Alt.2: Set B is a set of beam whose measurements are available as inputs of the AI/ML model (for prediction of set A)
* FL0: This is a good catch. This will be discussed in 9.2.3.2
  + Proposal 3: For BM-Case2, consider the scenario in which the UE orientation changes as a function of UE trajectory.
    - FFS: details of this function
* FL0: lack of discussion on “FFS part” for further discussion in this meeting on UE orientation.

|  |
| --- |
| **Agreement**   * For UE trajectory model, UE orientation can be independent from UE moving trajectory model. FFS on the details.   + Other UE orientation model is not precluded. |

#### FL2: (closed)Other assumptions

**Please indicate any other assumptions needs to be discussed and agreed in this meeting**

|  |  |
| --- | --- |
| Company | Comments |
| Samsung | We think the following proposal can be agreed for remaining assumption for LLS.   * + **Data allocation:**     - 8 RBs as baseline, companies can report larger number of RBs     - First 2 OFDM symbols for PDCCH, and following 12 OFDM symbols for data channel   + **Channel model:**     - LOS channel: CDL-D/E extension,     - NLOS channel: CDL-A/B/C extension,     - CDL-D extension, DS = 100ns as baseline.     - Companies explains details of extension methodology considering spatial consistency.     - Other channel models and DSs are not precluded. |
| FL3: | No urgency to have this proposal. Let’s close the discussion. |
|  |  |
|  |  |

# KPIs on AI/ML in beam management

## 2.1 Beam prediction accuracy related KPIs

Some observations/proposals were made in the contributions on beam prediction accuracy related KPIs:

**General discussion**

* Huawei [2]
  + Proposal 14: Since the prediction accuracy obtained from the AI/ML increases significantly with a larger K and then clearly outperforms the legacy baseline, adopt Top-K, K>1 (e.g., K=3, 5) for evaluation of spatial beam prediction accuracy.
  + Proposal 18: For temporal beam prediction evaluation, results for Top-K, K>1 should be presented in addition to Top-1 results.
    - The Top-1 predicted beam can be derived as the eventual result after the second round sweeping based on the AI/ML inferred Top-K beams.
  + FL0: the values of K can be reported by companies. Currently, K =3 and 5 are widely used.
* Interdigital [6]
  + Proposal 3: Support beam information related KPIs as optional for temporal measures.
    - Support average L1-RSRP difference of Top-1 predicted beam.
    - Support beam prediction accuracy (%) with multiple candidate margins (including 1 dB and other possible values) for Top-1 beam.
  + FL0: other margin can be reported by companies. There is no need to further agree on other KPIs.
* China Telecom [7]
  + Proposal 1: To evaluate the performance of AI/ML in beam management, at least following KPI should be considered as baseline, other options are not precluded:
    - Beam prediction accuracy (%) for Top-1 and/or Top-K beams.
    - The beam prediction accuracy (%) is the percentage of “the Top-1 predicted beam is one of the Top-K genie-aided beams”
    - Average L1-RSRP difference of Top-1 predicted beam
    - CDF of L1-RSRP difference for Top-1 predicted beam
  + FL0: no urgent need for down selection
* OPPO [8]
  + Proposal 6: Study another definition of L1-RSRP difference of Top-1 predicted beam
    - The difference between the predicted L1-RSRP of Top-1 predicted beam and the ideal L1-RSRP of the Top-1 genie-aided beam
  + FL0: There is no new/sufficient discussion compared with previous meeting for this proposal. Suggest to hold the discussion in later meetings
* Apple [21]
  + Proposal 1: The KPI for AI based beam prediction could be the beam prediction accuracy and the L1-RSRP distribution for the AI predicted beam. The KPI with RSRP can be used for making decision/drawing conclusion in the whole Rel-18 study item.

FL0: There is no intention to have down selection on the agreed KPIs in this meeting.

**Definition of beam prediction accuracy (%) if Top-1/K beams**

* Huawei [2]
  + Proposal 2: As KPI for the evaluation of the prediction accuracy, Option 2 should be selected, i.e., the beam prediction accuracy (%) is the percentage of “the Top-1 genie-aided beam is one of the Top-K predicted beams”.
* Spreadtrum [4]
  + Proposal 2: To evaluate the performance of AI/ML in beam management, Option 2 should be considered.
* Vivo [5]
  + Proposal 6: Support Option 2, i.e. the beam prediction accuracy (%) is the percentage of “the Top-1 genie-aided beam is one of the Top-K predicted beams”, to be the definition of beam prediction accuracy (%) for Top-1 and/or Top-K beams.
* China Telecom [7]
  + Proposal 1: To evaluate the performance of AI/ML in beam management, at least following KPI should be considered as baseline, other options are not precluded:
    - The beam prediction accuracy (%) is the percentage of “the Top-1 predicted beam is one of the Top-K genie-aided beams” (Note by FL0: option 2)
* OPPO [8]
  + Proposal 5: For beam prediction accuracy, adopt Option 2 (beam prediction accuracy (%) is the percentage of “the Top-1 genie-aided beam is one of the Top-K predicted beams”) for AI/ML beam prediction.
  + Proposal 6: Study another definition of L1-RSRP difference of Top-1 predicted beam
    - The difference between the predicted L1-RSRP of Top-1 predicted beam and the ideal L1-RSRP of the Top-1 genie-aided beam
* CATT [12]
  + Proposal 1: To evaluate the performance of AI/ML in beam management, the definition of beam prediction accuracy for Top-1 and/or Top-K beams is Option 2, i.e., the Top-1 genie-aided beam is one of the Top-K predicted beams.
* Xiaomi [17]
  + Proposal 1: For the definition of beam prediction accuracy (%) for Top-1 and/or Top-K beams, Option 2 is preferred.
* CMCC [18]
  + Proposal 1: The definition of beam prediction accuracy (%) for Top-1 and/or Top-K beams is:
    - Option 2: The beam prediction accuracy (%) is the percentage of “the Top-1 genie-aided beam is one of the Top-K predicted beams”
  + Proposal 2: The definition of beam prediction accuracy (%) with 1 dB margin for Top-K beams is:
    - The percentage of “the ideal highest L1-RSRP of the predicted Top-K beams is within 1 dB of the L1-RSRP of the Top-1 genie-aided beam”

Based on the above proposals, the following proposals are proposed:

#### FL3: Definition on beam prediction accuracy (%) for Top-1/K beams

**Proposal 2-1-1a:**

* **The definition of beam prediction accuracy (%) for Top-1 and/or Top-K beams:**
  + **Option 1 (optional): The beam prediction accuracy (%) is the percentage of “the Top-1 predicted beam is one of the Top-K genie-aided beams”**
  + **Option 2 (baseline): The beam prediction accuracy (%) is the percentage of “the Top-1 genie-aided beam is one of the Top-K predicted beams”**

|  |  |
| --- | --- |
| Supporting companies | Futurewei, Google, MediaTek, LG, Xiaomi, HwHiSi, OPPO, Spreadtrum, NTT DOCOMO, Lenovo, Ericsson, vivo, CATT, ZTE, NVIDIA |
| Objecting companies |  |

**Please provide your view Proposal 2-1-1a, if any.**

|  |  |  |
| --- | --- | --- |
| Company | Comments | |
| OPPO | Support.  Hope there could be further down selection for comparison. If so, our preference is to take Option 2 as baseline. | |
| Nokia | Support | |
| Lenovo | We prefer Option 1. | |
| Company | Yes/No | Comments |
| FL2 |  | As suggested by MediaTek, please continue comment on the following proposals. If we separate the definition, maybe no need to set baseline for now, as other KPIs.  **Proposal 2-1-1b:**   * **The definitions of beam prediction accuracy (%):**   + **The beam prediction accuracy (%) for Top-K beams () is the percentage of “the Top-1 genie-aided beam is one of the Top-K predicted beams”**   + **The beam prediction accuracy (%) for Top-K genie-aided beams is the percentage of “the Top-1 predicted beam is one of the Top-K genie-aided beams”** |
| HW/HiSi | No | We prefer the original proposal 2-1-1a which has large support and we would be supportive to use Option 2 as baseline.  For the updated proposal, we do not agree with the second bullet, since it would overestimate the accuracy. This definition may not guarantee that the Top-1 genie-aided beam which is the best global beam would be correctly predicted. Also, capturing other suboptimal genie-aided beams (other than the Top-1 genie-aided beam) can be misleading since they may have a large performance gap to the Top-1 genie-aided beam  We suggest to the following update:  **Updated Proposal 2-1-1b:**   * **The definitions of beam prediction accuracy (%):**   + **(baseline) The beam prediction accuracy (%) for Top-K beams () is the percentage of “the Top-1 genie-aided beam is one of the Top-K predicted beams”**   + **(optional) The beam prediction accuracy (%) for Top-K genie-aided beams is the percentage of “the Top-1 predicted beam is one of the Top-K genie-aided beams”** |
| vivo | No | We still think it is better to have a baseline for further evaluation. We prefer the suggestion from Mr. Chair during online discussion, which would avoid lengthy discussion.  **Updated Proposal 2-1-1b:**   * **The options to evaluate beam prediction accuracy (%) ~~for Top-1 and/or Top-K beams~~:**   + **Option 1 (optional): The beam prediction accuracy (%) is the percentage of “the Top-1 predicted beam is one of the Top-K genie-aided beams”**   + **Option 2 (baseline): The beam prediction accuracy (%) is the percentage of “the Top-1 genie-aided beam is one of the Top-K predicted beams”** |
| Spreadtrum | No | Agree with VIVO, we need a baseline for comparison. To clarify whether Top-K beams refers to predicted or genie-aided beam, we think change Top-K beams to Top-K predicted beam is better.  **Updated Proposal 2-1-1b:**   * **The definitions of beam prediction accuracy (%):**   + **(baseline) The beam prediction accuracy (%) for Top-K predicted beams () is the percentage of “the Top-1 genie-aided beam is one of the Top-K predicted beams”**   + **(optional) The beam prediction accuracy (%) for Top-K genie-aided beams is the percentage of “the Top-1 predicted beam is one of the Top-K genie-aided beams”** |
| CEWiT | No | We agree with Vivo, we need to have a baseline for comparison. |
| NTT DOCOMO | No | We prefer to have the baseline of the beam prediction accuracy evaluation for the calibration purposes. Also, option2 in the original proposal seems sufficient as the definition of beam prediction accuracy. |
| OPPO | No | In our reading, the beam prediction accuracy is widely used performance metric. It would be great to determine the baseline KPI as early as possible for comparison reasons.  The original wording on beam prediction accuracy causes no confusion to us. But to address the concern raised during GTW, the rewording from vivo looks good. |
| LG | No | We agree with vivo. |
| CMCC | No | We think Spreadtrum’ s version is more clear. VIVO’s update is also fine to us. |
| Fujitsu | N | Prefer the original proposal 2-1-1a |
| ZTE | No | We support vivo’s update. |
| FL2 |  | |  | | --- | | Agreement   * To evaluate the performance of AI/ML in beam management, further study the following KPI options:   + Beam prediction accuracy related KPIs, may include the following options:     - Beam prediction accuracy (%) for Top-1 and/or Top-K beams, FFS the definition:       * Option 1: The beam prediction accuracy (%) is the percentage of “the Top-1 predicted beam is one of the Top-K genie-aided beams”       * Option 2: The beam prediction accuracy (%) is the percentage of “the Top-1 genie-aided beam is one of the Top-K predicted beams”   …… |   Based on the current discussion and above agreements in RAN 1 #109e, I think Beam prediction accuracy (%) for Top-1 beam is the most important one. For Top-K predicted beam has the best beam, or the Top-1 prediction beam in one of K best beams, the first one has more support. Therefore, maybe we can clear up as the following three metrics, which might be easier for results collection.  Although I think we don’t have baseline for any of the KPIs, making Top-1/K(%) as optional may be acceptable.  Please share your views on **Proposal 2-1-1b** and/or **Proposal 2-1-1c**, or if you strongly want to keep or remove “Top-1/K(%)”.  **Proposal 2-1-1c:**   * Beam prediction accuracy (%) for Top-1 and/or Top-K beams”, consider the following options:   + Top-1 (%): the percentage of “the Top-1 genie-aided beam is Top-1 predicted beam”   + Top-K/1 (%): the percentage of “the Top-1 genie-aided beam is Top-K predicted beam”   + Top-1/K (%) (Optional): the percentage of “the Top-1 predicted beam is one of the Top-K genie-aided beams” |
| Lenovo |  | * Option 1 (in proposal 2-1-1a) rewards a scheme whose predicted beam is one of the top K beams with the largest gains, whereas Option 2 (in proposal 2-1-1a) may reward a scheme that predicts the best beam + worst K-1 beams. We therefore *prefer* Option 1 in proposal 2-1-1a which provides a more robust metric. * We can accept Proposal 2-1-1b for the sake of progress. * Regarding Proposal 2-1-1c: We are fine with the first bullet. Among 2nd and 3rd bullets, we prefer the 3rd bullet for the same reason explained above. We want to keep “Top-1/K(%)” bullet. |
| CATT |  | We support Proposal 2-1-1c.  Just one typo in the 2nd sub-bullet of Proposal 2-1-1c: Top-K/1 (%): the percentage of “the Top-1 genie-aided beam is one of the Top-K predicted beam” |
| MediaTek |  | Since most of companies prefer to keep baseline or optional, we can support the updated proposal 2-1-1b by Spreadtrum. Removed unnecessary parenthesis from Spreadtrum’s version. We think Top-1 is the subset of Top-K. We don’t have to separately define it as in Proposal 2-1-1c.  **Updated Proposal 2-1-1b:**   * **The definitions of beam prediction accuracy (%):**   + **(baseline) The beam prediction accuracy (%) for Top-K predicted beams ~~()~~ is the percentage of “the Top-1 genie-aided beam is one of the Top-K predicted beams”**   + **(optional) The beam prediction accuracy (%) for Top-K genie-aided beams is the percentage of “the Top-1 predicted beam is one of the Top-K genie-aided beams”** |
| Futurewei |  | We don’t think there is need to further separate Top-1 and Top-K/1 as specified in Proposal 2-1-1c. We prefer the original proposal with the wording modification from vivo.  FL2: In the result collection, Top 1 and Top K =1, 3, 5 etc will be separated reported, unless you think Top-1(%) is useless. |
| Intel |  | We think only two options are enough. For Proposal 2-1-1b, the first sub-bullet can be removed. The second sub-bullet should be the baseline and the third sub-bullet should be optional.  FL2: In the result collection, Top 1 and Top K =1, 3, 5 etc will be separated reported, unless you think Top-1(%) is useless. |
| CAICT |  | We can support proposal 2-1-1c and also fine to add baseline for option 1 and 2. |
| vivo |  | To keep the original two options seems better, and we are also OK for Proposal 2-1-1c. |
| ZTE |  | We support Proposal 2-1-1c, where both top-1 and top-K beam prediction accuracy are appropriately defined. Proposal 2-1-1b only includes the two definitions of top-K beam prediction accuracy. One little comment based on CATT's update: Top-K/1 (%): the percentage of “the Top-1 genie-aided beam is one of the Top-K predicted beams”.  FL2 : will be considered in next updates |
| FL2 |  | The motivation to separate Top-1(%) from Top-K (%) is because Top-1 (%) will be separated in the end for results collection. And it is the same for either option.  Please consider Proposal 2-1-1c and provide comment again. |
| Samsung |  | We are fine with Proposal 2-1-1c. |
| Xiaomi |  | We are OK with the proposal 2-1-1c. There is a typo in the 2nd sub-bullet:   * + Top-K/1 (%): the percentage of “the Top-1 genie-aided beam is one of the Top-K predicted beam” |
| Ericsson |  | Ok with Proposal 2-1-1c. |
| Qualcomm |  | As Top-1 accuracy is a special case of Top-K accuracy, we do not see the need to separate them out as in Proposal 2-1-1c. We prefer the Updated Proposal 2-1-1b by VIVO. |
| Fujitsu |  | Share the view of Futurewei. For proposal 2-1-1c, what’s the difference on option 1 and option 2/3 with K=1 for Top-K? If there is no difference, the option 1 is just a special value of K for option 2/3  FL3: as explained to Intel, K=1 and >1 will be reported separately in the end. Actually in my understanding, K=1 can be the baseline if we targets to obtain the best Tx beam for DL transmission. |
| LG |  | Fine with Proposal 2-1-1c. |
| NTT DOCOMO |  | We prefer proposal 2-1-1b for the simplicity. Also as the majority supports one definition as the baseline for the calibration purpose, we think the description should be reflected based on the opinions. |
| Spreadtrum |  | We support Proposal 2-1-1c, and agree with ZTE’s update. |
| HW/HiSi |  | We have a question for clarification: would proposal 2-1-1c mean that all options should be reported or only one of them?  If companies can choose freely across the options, it will be very difficult to compare results. Therefore we think at least one common option should be supported as baseline. We propose the option that has had most support.  For the last bullet, we think it is not really a useful metric, at least when considering 2nd round beam sweeping, since it is overestimating the performance too much, since there could be a large RSRP difference between the best and second-best beam.  **Updated Proposal 2-1-1c:**   * Beam prediction accuracy (%) for Top-1 and/or Top-K beams”, consider the following options:   + Top-1 (%): the percentage of “the Top-1 genie-aided beam is Top-1 predicted beam”   + (baseline) Top-K/1 (%): the percentage of “the Top-1 genie-aided beam is Top-K predicted beam”   Top-1/K (%) (Optional): the percentage of “the Top-1 predicted beam is one of the Top-K genie-aided beams” |
| CMCC |  | We prefer the version of proposal 2-1-1b. HW’s update is also fine. |
| FL3 |  | @Huawei, all the KPIs are not mandated to report so far. That’s the reason why I didn’t mark baseline to avoid confusion.  @ All, in results collection, we will have separate results for K=1 and K=3, 5, etc. Several companies reported the result with “Top-1 %”, “Top-3 % opt 1” “Top-3 % opt 2” already  K=1 is special and can be the baseline if the goal of AI for BM for some sub-case is to obtain the best Tx beam for DL transmission. I think separate it out is better.  The purpose for this proposal is for evaluation in this meeting. Whether do down selection, it can be discussed further.  Encourage companies to provide views on **Proposal 2-1-1d:**  **Proposal 2-1-1d:**   * The options to evaluate beam prediction accuracy (%):   + Top-1 (%): the percentage of “the Top-1 genie-aided beam is Top-1 predicted beam”   + Top-K/1 (%): the percentage of “the Top-1 genie-aided beam is one of the Top-K predicted beams”   + Top-1/K (%) (Optional): the percentage of “the Top-1 predicted beam is one of the Top-K genie-aided beams”   + Where K =3, 5, and other values can be reported by companies. |
| NVIDIA |  | Proposal 2-1-1d is fine. Better example values for K may be (2, 4) than (3, 5) |
| Xiaomi |  | Support proposal 2-1-1d and share same view as NVIDIA that (2,4) is better than (3,5) since the number of beams in legacy L1 beam report is {1,2,4}. |
| ZTE |  | We are fine with the latest proposal. Besides, K=4 can be added as a candidate since it has been used by many companies. |
| OPPO |  | We are generally fine the definition of beam prediction accuracy.  But as others, we also think K = 2 and 4 are more common numbers for beam prediction which is aligned with that of beam reporting per instance. |
| HW/HiSi |  | We are fine with the proposal. Open to add other values if requested from other companies. But for comparison, we should at least have 2 common values, we think 3 and 5 are fine. |
| CATT |  | We are fine with the proposal 2-1-1d. For the value of K, suggest delete the example, just say “Where the values of K can be reported by companies”. Perhaps this can avoid the discussion of value. |
| Samsung |  | We support Proposal 2-1-1d. Also, we think ‘K = 2, 3, 4’ is better than ‘K = 3, 5’, for finer granularity of K. |
| LG |  | We are fine with the proposal. And open for other values. |
| Spreadtrum |  | We are fine with the proposal. And open for other values. |
| Lenovo | Yes | We are OK with the proposal. |
| Ericsson |  | Support proposal 2-1-1d, agree with Nvidia’s comment on the values of K |
| MediaTek |  | We still prefer our updated Proposal 2-1-1b because we think it is not necessary to define Top-1 (%) separately, but we can live with Proposal 2-1-1d. |

**Clarification on “Top1 genie-aided Tx beam”**

* Huawei [2]
  + Proposal 3: For DL Tx beam prediction, the Top-1 genie-aided Tx beam is defined as the Tx beam that results in the largest RSRP at the UE side
    - For Case A (L1-RSRP of Tx beams in Set B, measured by a “best” Rx beam), the Top-1 genie-aided Tx beam should be the Tx beam ID that results in the largest RSRP over all Tx and Rx beams
    - For Case B (L1-RSRP of Tx beams in Set B, measured by the same Rx beam), the Top-1 genie-aided TX beam should be the Tx beam ID that results in the largest RSRP over all Tx beams with that specific Rx beam

#### FL3: Clarification on Top1 genie-aided Tx beam

**Proposal 2-1-2a:**

* **For DL Tx beam prediction, the Top-1 genie-aided Tx beam is defined as the Tx beam that results in the largest L1-RSRP, FFS:** 
  + **Option A, the Top-1 genie-aided Tx beam is the Tx beam ~~ID~~ that results in the largest L1-RSRP over all Tx and Rx beams**
  + **Option B, the Top-1 genie-aided TX beam is the Tx beam ~~ID~~ that results in the largest L1-RSRP over all Tx beams with specific Rx beam(s)**

**Please provide your view Proposal 2-1-2a, if any.**

|  |  |  |  |
| --- | --- | --- | --- |
| Company | Comments | | |
| Google | Support | | |
| FL1 | Deleted “ID” in proposal 2-1-2a | | |
| Xiaomi | support | | |
| OPPO | Fine with preference on Option B.  As this is for DL Tx beam prediction, if Option A is taken, then we tend to understand it as DL Tx-Rx beam prediction. | | |
| Spreadtrum | Support | | |
| NTT DOCOMO | Support the proposal | | |
| vivo | Support.  Between Opt A and Opt B, we support Opt A to ensure fair comparison between Tx-Rx beam pair prediction and Tx beam prediction. | | |
| Samsung | Support | | |
| CATT | Support | | |
| FL1 | Please continue to provide comments and provide your preference if any. | | |
| Company | Y/N | Opt A or B | Comments |
| HW/HiSi |  |  | We think the FFS can be removed and both options can be reported depending on the functionality of the AI model. For Tx-Rx beam pair prediction, Option A is more suitable and for Tx beam prediction, Option B would be more suitable. |
| nvidia |  |  | It may be helpful to first discuss UE Rx beam determination, the discussion of which may impact on the selection here. |
| Futurewei | Y |  | Companies can decide which option to use. |
| Apple |  |  | Both Option A and Option B can be used, sharing a similar view as Huawei here. |
| vivo |  |  | We are okay to keep both Opt A and Opt B. |
| Spreadtrum |  |  | In our opinion, the choice of option A or B depends on the type of AI beam prediction. If it is beam pair prediction, option 1 should be used, and if it is Tx beam prediction, option 2 should be selected. So both options should remain |
| CEWiT |  |  | We are okay with keeping both the options at this stage. |
| CAICT |  |  | Same view as HW. |
| Samsung |  |  | We think companies can report both options depending on their assumption on Rx beam selection. Regarding DL Tx-Rx beam pair prediction, we think other definition may be needed. Therefore, we would like to clarify the proposal as follows:  **=> Proposal 2-1-2a:**   * **For DL Tx beam prediction, the Top-1 genie-aided Tx beam is defined as the Tx beam that results in the largest L1-RSRP, FFS:**    + **Option A, the Top-1 genie-aided Tx beam is the Tx beam ~~ID~~ that results in the largest L1-RSRP over all Tx and Rx beams**   + **Option B, the Top-1 genie-aided TX beam is the Tx beam ~~ID~~ that results in the largest L1-RSRP over all Tx beams with specific Rx beam(s)**   **For DL Tx-Rx beam pair prediction, the Top-1 genie-aided Tx beam is defined as the Tx beam associated with Tx-Rx beam pair that results in the largest L1-RSRP.** |
| NTT DOCOMO |  |  | Can leave both options at this stage. |
| OPPO |  | Option B | In the main bullet, it says “for DL Tx beam prediction”, so we think for fair comparison, the genie-aided beam should be selected by using specific Rx beam(s), rather than sweeping all Rx beams. But if it comes to DL Tx-Rx beam pair prediction, we think Option A should be adopted. |
| Nokia | Y |  | Let’s keep both open. |
| LG |  |  | We can live with both options. |
| CMCC | Y |  | Support the proposal. Whether adopt option A or Option B can be reported by companies. |
| Fujitsu |  |  | Share the similar view with Nvidia. For DL Tx beam prediction, it’s suggested to assume one wide Rx beam in UE to avoid the selection on Rx beam. |
| ZTE |  |  | For DL Tx beam prediction, it should be clarified how to choose the Rx beam for measurement. We think option A corresponds to DL Tx-Rx beam pair prediction since both Tx beam and Rx beam need to be predicted. Option B is suitable for DL Tx beam prediction where Rx beam is fixed for measurement. |
| FL2 |  |  | Majority companies support to keep both of options. However, similar as the comments online on other definition, we’d better to keep one definition for one term. Therefore, I updated the proposal for “potential down selection”.  Based on Samsung’s comment, Option A seems more suitable for Tx-Rx beam pair prediction. Therefore, I draft a separate proposal for it.  please provide your views on **Proposal 2-1-2b** and **Proposal 2-1-3a**.  **Proposal 2-1-2b:**   * **For DL Tx beam prediction, the definition of Top-1 genie-aided Tx beam ~~is defined as the Tx beam that results in the largest L1-RSRP,~~ considers the following options for potential down selection: ~~FFS:~~**    + **Option A, the Top-1 genie-aided Tx beam is the Tx beam that results in the largest L1-RSRP over all Tx and Rx beams**   + **Option B, the Top-1 genie-aided Tx beam is the Tx beam that results in the largest L1-RSRP over all Tx beams with specific Rx beam(s)**   **Proposal 2-1-3a:**   * **For DL Tx-Rx beam pair prediction, the definition of Top-1 genie-aided Tx beam, is the Tx beam that results in the largest L1-RSRP over all Tx and Rx beams** |
| Company | Y/N  For 2b | Y/N  For 3a | Comments |
| Lenovo | Y | Y | In Proposal 2-1-2b, we prefer Option B. |
| CATT | Y | Y | In Proposal 2-1-2b, we also prefer Option B.  Support proposal 2-1-3a. |
| MediaTek | Y | Y | Support proposal 2-1-2b and prefer Option A.  Support proposal 2-1-3a. |
| Futurewei | Y | Y | For Proposal 2-1-2b, we agree with Lenovo and CATT that Option B is more suitable for DL Tx prediction. |
| Intel | N | Y | For 2b, we don’t think down-selection is necessary since both options can provide useful information and companies can report assumptions. In this case, we can just keep both Options. We understand FL’s motivation for down-selection, but this issue is different from definition of the metric itself. Additionally, of the two options listed, we prefer Option B.  Ok with 3a.  FL2: Just for “potential down selection”, if companies think it is helpful to draw conclusion. We can keep both. |
| CAICT | Y | Y | In Proposal 2-1-2b, Option B is preferred. |
| Apple |  |  | for 2-1-2b, we support Option B. for 2-1-3a, can we have an analogous formulation as 2-1-2b?  **Option A: For DL Tx-Rx beam pair prediction, the definition of Top-1 genie-aided Tx beam, is the Tx beam that results in the largest L1-RSRP over all Tx and Rx beams**  **Option B: For DL Tx-Rx beam pair prediction, the definition of Top-1 genie-aided Tx beam, is the Tx beam that results in the largest L1-RSRP over all Tx over all Tx beams with specific Rx beam(s)**  FL2: I think it doesn’t make sense to only evaluate for given Rx beam(s). But we can hear more views from other companies. |
| vivo |  |  | OK with the two proposals from FL. Prefer Opt A in 2-1-2b. |
| ZTE |  |  | We prefer option B in Proposal 2-1-2b. Further down selection depends on whether Rx beam is fixed or not during the measurement process. |
| Samsung | Y | Y | Regarding Proposal 2-1-2b, we prefer to keep both options in this meeting |
| Xiaomi |  | Y | For Option B in proposal 2-1-2b, we want to clarify that “specific Rx beam(s)”. If there are 2 specific Rx beams, the number of Top-1 genie-aided Tx beam is 1 for 2 Rx beams, or 1 for each Rx beam? We are not clear about the typical use case for more than 1 Rx beam. We prefer to remove “(s)” in Option B.  And Option A and Option B may have different use case, we suggest to keep both of them. |
| Ericsson | Y | Y | Ok with the updated FL proposals |
| Qualcomm | Y | Y | In Proposal 2-1-2b, we prefer to keep both options for this meeting. |
| Fujitsu | Y | Y | In Proposal 2-1-2b, we prefer Option B. but how to select the specific Rx beam is still not clear. |
| LG | Y | Y | Ok with the updated FL proposals |
| NTT DOCOMO | Y | Y | For down selection in Proposal 2-1-2b, option B is preferred. |
| Spreadtrum | Y | Y | In Proposal 2-1-2b, we prefer Option B.  Support proposal 2-1-3a. |
| HW/HiSi | Y | Y | We support both proposals:  As a comment on 2-1-2b, since the FFS has been removed, both options can coexist. There is no need to further down-selection between them. |
| FL3 |  |  | Hope these two proposals are stable.  **Proposal 2-1-2c:**   * **For DL Tx beam prediction, the definition of Top-1 genie-aided Tx beam considers the following options for potential down selection:**    + **Option A, the Top-1 genie-aided Tx beam is the Tx beam that results in the largest L1-RSRP over all Tx and Rx beams**   + **Option B, the Top-1 genie-aided Tx beam is the Tx beam that results in the largest L1-RSRP over all Tx beams with specific Rx beam(s)**     - **FFS on specific Rx beam(s)**   **Proposal 2-1-3a:**   * **For DL Tx-Rx beam pair prediction, the definition of Top-1 genie-aided Tx beam, is the Tx beam that results in the largest L1-RSRP over all Tx and Rx beams** |
| NVIDIA |  |  | Support Proposals 2-1-2c and 2-1-3a. |
| Futurewei |  |  | We are ok with Proposals 2-1-2c and 2-1-3a. |
| Xiaomi |  |  | Support proposal 2-1-3a and fine with proposal 2-1-2c |
| ZTE |  |  | We are fine with Proposals 2-1-2c and 2-1-3a. There is a typo in Proposal 2-1-3a that the second comma can be deleted. |
| OPPO |  |  | Support Proposal 2-1-3a.  On Proposal 2-1-2c: we are fine to the further down selection. For Option A, we would like to note that it’s for DL Tx beam prediction, assuming specific Rx beam(s) pre-determined (either fixed or the best one). If the genie-aided Tx beam is selected by sweeping all Rx beam, then it seems not fair to calculate the beam prediction accuracy %. We prefer Option B, if down selected. |
| HW/HiSi |  |  | We would like to remove the “potential down-selection” for **Proposal 2-1-2c: .**  The reason is that the determination of the RX beam can also be done in the legacy P3 after second round beam sweeping. Since if the AI model is used to replace the P1 legacy procedure, where RX beam determination is done in P3, it is not meaningful to use Option A. Both options should co-exist and can be reported. |
| CATT |  |  | Support Proposal 2-1-3a.  For Proposal 2-1-2c, we share the same view with OPPO. We also prefer Option B, if down selected. |
| Samsung |  |  | Support Proposal 2-1-2c and Proposal 2-1-3a. |
| LG |  |  | Support Proposals 2-1-2c and 2-1-3a. |
| Spreadtrum |  |  | Support Proposals 2-1-2c and 2-1-3a. And we prefer Option B in Proposal 2-1-2c. For Option A, if the Top-1 Tx beam is chose from all Tx and Rx beams, this beam is actually an optimal beam pair and not the Tx beam. |
| Lenovo |  |  | OK with both the proposals. |
| Ericsson |  |  | Support Proposals 2-1-2c and 2-1-3a. |
| MediaTek |  |  | We are ok with Proposal 2-1-2c and Proposal 2-1-3a. |

**Other aspects**

* Futurewei [1]
  + Observation 4: When evaluating AI/ML model performance, using “Average L1-RSRP difference of Top-1 (or Top-K) predicted beam” alone may not directly indicate the performance unless the average L1-RSRP difference between the ideal L1-RSRP of the Top-1 genie-aided beam and the ideal L1-RSRP of the Top-K genie-aided beams in the (testing) dataset is known.
  + Proposal 2: For AI/ML based spatial beam prediction, to help performance evaluation discussion, companies are encouraged to share simulation details for the dataset generation and provide the average L1-RSRP difference between the ideal L1-RSRP of the Top-1 genie-aided beam and the ideal L1-RSRP of the Top-K genie-aided beams in the training/testing dataset.
* Ericsson [11]
  + Observation 1: The agreed simulation scenarios might have heavily skewed beam statistics. AI/ML models can be trained to work well for common beams and ignore uncommon beams. The poor performance of AI/ML models on uncommon beams might not be reflected in average beam prediction statistics. Visualizing the edge percentiles of the L1-RSRP CDF could be one method to illustrate the ability to predict uncommon beams
* Qualcomm [26]
  + Proposal 4: For BM-Case1 and BM-Case2, study the impact of incorporating beam prediction quality information (e.g., a measure for prediction confidence such as std of predicted RSRPs) on evaluating the performance of AI/ML model, using the agreed KPIs
    - Note: The results from this study could help in defining criteria or metrics for AI/ML model performance monitoring which could lead to model activation/deactivation or updating of AI/ML models.

Other than the points raised by companies, by reading companies simulation results, FL feels that directly using average L1-RSRP different of Top-1/K beam may have some issues (e.g. may not be applicable), especially for generalization performance when basic configuration changed.

#### FL2: (closed)Other aspects for L1-RSRP related KPIs

**Please indicate whether any other aspects including additional KPIs/definitions are needed for L1-RSRP difference? And please explain the reason, at least including:**

* **A1:** For AI/ML based spatial beam prediction, to help performance evaluation discussion, companies are encouraged to share simulation details for the dataset generation and provide the average L1-RSRP difference between the ideal L1-RSRP of the Top-1 genie-aided beam and the ideal L1-RSRP of the Top-K genie-aided beams in the training/testing dataset. Futurewei [1]
* **A2:** Visualizing the edge percentiles of the L1-RSRP CDF could be one method to illustrate the ability to predict uncommon beams. Ericsson [11]
* **A3:** For BM-Case1 and BM-Case2, study the impact of incorporating beam prediction quality information (e.g., a measure for prediction confidence such as std of predicted RSRPs) on evaluating the performance of AI/ML model, using the agreed KPIs. Qualcomm [26]
* **A4:** Other comments

|  |  |
| --- | --- |
| Company | Comments |
| Google | We are open to consider other aspects, but we do not quite understand what A2 means. Could proponent clarify it a bit?  FL1: @Ericsson can help to clarify a little bit. |
| FL1 | Proponents can further explain the motivations and provide wording for a proposal, if needed. . |
| Ericsson | We have observed that in some scenarios, baseline performs better than AI/ML method for the edge percentiles in the CDF, hence it is important the CDF clearly illustrates edge percentiles. However, given the agreement to present results for the 5th percentile in system level simulations, we don’t think there is a need to have any agreement on the above. |
| HW/HiSi | We think this question should be given lower priority for this meeting. |
| Futurewei | We think providing L1-RSRP difference between the ideal L1-RSRP of the Top-1 genie-aided beam and the ideal L1-RSRP of the Top-K genie-aided beams in the testing dataset is very helpful in understanding companies’ results. For example, if the true L1-RSRP differences between Top-8 beams are all very close in the dataset used in testing, e.g., within 1 dB in, then a prediction performance of 1dB for L1-RSRP difference of Top-1 predicted beam does not necessarily mean the performance is good. Thus, we strongly encourage company to provide such information in their results. |
| Nokia | Adding some percentiles of L1-RSRP error (e.g. 95%, 90%, 80%, 50%) to be analyzed​/reported can help to better evaluate the ML model performance for some aspects (e.g. generalization). The average L1-RSRP error may not be sufficient. |
| MediaTek | Can be deprioritized, we already have a lot of KPIs. Companies are welcome to include additional KPIs in the tdoc for the corresponding evaluation findings. |
| Intel | While we are OK with lower priority for this issue, we think the CDF of L1-RSRP difference can provide good information about performance of the AI/ML model. |
| CATT | We also think this issue can be deprioritized in this meeting. |
| LG | This issue can be deprioritized. |
| FL3 | Companies are encouraged to study the aspects mentioned here.  The discussion in this meeting is closed. |

## 2.2 System performance related KPIs

### 2.2.1 (Closed) User throughput

Other than beam measurement related KPIs, several companies mentioned that the system performance shall be also evaluated:

* Interdigital [6]:
  + Proposal 2: Support system performance related KPIs as mandatory KPIs.
    - Support Avg. and 5% UE tput for system performance KPIs.
  + Proposal 5: Prioritize system performance related KPIs and beam information related KPIs than other KPIs.
* Samsung [24]:
  + Proposal 7: Shannon capacity-based simplified model for UPT can be further considered as additional system performance related KPI.
* Qualcomm [26]
  + Proposal 7: At least for spatial domain beam prediction, consider spectral efficiency CDF for SLS evaluations as a KPI.

FL0: there is no need for further discussion on user throughput in this meeting.

### 2.2.1 RS overhead

There were several proposals/discussions related to RS overhead:

* Huawei/HiSi [2]:
  + Proposal 10: For the evaluation of the overhead for spatial domain AI/ML-based BM, two metrics should be reported:
    - The RS overhead, consisting of the beams being swept in Set B and the Top-K beams for P2 beam sweeping after inference (if applicable)
      * RS OH = N + K for K > 1 and RS OH = N for K = 1, where N is the number of beams in Set B and K is the number of Top-K selected beams.
    - The RS overhead reduction compared to an exhaustive beam sweep over set A
      * RS OH RD [%] = 1-(N+K)/M for K > 1 and RS OH [%] = 1-N/M for K =1, where N is the number of beams in Set B, K is the number of Top-K selected beams and M is the number of beams in Set A.
  + Proposal 13: For the evaluation of the overhead for temporal domain AM//ML-based BM, the observation and prediction window are jointly considered, and two metrics should be reported
    - The RS overhead, consisting of the beams being swept in Set B during observation and the Top-K beams for P2 beam sweeping during prediction (if applicable)
      * for K>1 and for K = 1
    - The RS overhead reduction compared to an exhaustive beam sweep over Set A during observation and the Top-K beams for P2 beam sweeping during prediction (if applicable)
      * for K > 1 and for K = 1
    - Where: M is beams in Set A, N is beams in Set B and K is the number of beams as inference output
* ZTE [3]
  + Proposal 2: RS overhead reduction should be considered as a basic KPI for evaluation and should be further studied with considering following factors: the number of UE, the beam pattern, and the refined beam sweeping procedure.
* Spreadtrum [4]:
  + Proposal 3: For RS overhead or RS overhead reduction, option 1 should be considered as KPI for spatial domain beam prediction.
* Vivo [5]
  + Proposal 7: The metric of beam sweeping overhead reduction is calculated as 1-N/M where N is the number of beams required for measurement in both non-AI algorithm and AI algorithm, and M can be the total number of all possible beams to be predicted.
* OPPO [8]
  + Proposal 7: For BM-Case1, study how to accurately capture the overhead, considering beam measurement on Set B and potential follow-up measurement after beam prediction.
  + Proposal 8: For BM-Case2, study how to capture the overhead reduction, considering the T1 duration (measurement on Set B) and T2 duration (prediction among Set A).
* Ericsson [11]
  + Proposal 3: Define a RS measurement overhead KPI, e.g. N/M where N is the number of beams measured by a UE, and M is the total number of beams.
* Fujitsu [13]
  + Proposal 3: For the KPI of RS overhead reduction, it is suggested to consider three alternatives of predicted beams.
    - Alt.1: DL Tx beam prediction
    - Alt.2: DL Rx beam prediction
    - Alt.3: Beam pair prediction
  + Proposal 4: Regarding the three alternatives of predicted beam, the KPI of RS overhead reduction is suggested to be calculated as：

|  |  |  |  |
| --- | --- | --- | --- |
|  | Tx beam prediction | Rx beam prediction | Beam pair prediction |
| RS overhead reduction |  |  |  |

* FL0: please check current wording. I think it can be covered by the case in your mind.
* Lenovo [15]
  + The RS overhead reduction, for at least top-1 spatial-domain beam prediction, is given by

,

* + where *N* is the number of beams (with reference signal, i.e., (SSB and/or CSI-RS)) required for measurement and *M* is the total number of beams.
  + *Note* that this metric is meaningful for reference signals and there would not be any RS overhead reduction when only SSBs are considered, because *M* SSBs would be transmitted on *M* beams.
  + To accommodate the AI/ML models that perform varying number of beam measurements in each time slot, the above metric may be modified as follows [4], [6]:

,

* + Where is the number of beam measurements in time slot and the is the total number of time slots.
  + Thus, the above metric is a general version of the first metric for RS overhead reduction.
* CAICT [16]:
  + Proposal 3: RS overhead calculation for DL tx beam prediction and temporal domain beam prediction (BM-Case2) could be reported by different companies.
* Xiaomi [17]
  + Proposal 3: Study the following options on RS overhead reduction for temporal beam prediction:
    - Option 1: "RS " OH[%]=1-N/(N+M)
      * For the case of the same periodicity of history measurement instance and future time instance, where N is the number of history measurement instance and M is the number of predicted future time instance.
    - Option 2: "RS " OH[%]=1-1/L
      * For the case of the periodicity of history measurement instance is L times of that of future time instance.
* Nokia [19]:
  + RS overhead reduction at least for spatial-domain beam prediction at least for Top-1 beam

|  |
| --- |
| where N is the number of beams (with reference signal (SSB and/or CSI-RS)) required for measurement, M is the total number of beams. Non-AI/ML approach based on the measurement of these M beams may be used as a baseline.  When N is variable, the overhead reduction is computed using an average measurement set size, such that  Where is the number of beams required for measurement during time slot |

* Samsung [24]
  + Proposal # 6: For RS overhead reduction, further study the following options:
    - Option 1: at least for BM-Case 1
      * where N is the number of beams (pairs) (with reference signal (SSB and/or CSI-RS)) required for measurement (in Set B and in Top-K beams (pairs) for P2 beam sweep after inference (if applicable))
      * where M is the total number of beams (pairs) to be predicted (in Set A)
    - Option 2: at least for BM-Case 2
      * Where N\_n is the number of beams (pair) (in Set B and in Top-K beams (pairs) for P2 beam sweep after inference (if applicable)) required for measurement during time slot n
      * where M is the total number of beams (pair) to be predicted (in Set A)
    - FFS on other options
* DoCoMo [25]:
  + For example, the following equation can be considered as KPI for RS overhead reduction.
  + Observation 1: Additional beam measurements might be necessary for PDSCH/PDCCH reception with top1/K predicted beam(s), when the top-1/K predicted beam(s) are not included in beams measured for the beam prediction.
  + Proposal 1: Discuss the requirement of actual QCL relation, and consider the additional RS measurement overhead to obtain the actual QCL relation if necessary.



Figure 1. Additional beam measurements with the top-1 predicted beam for reception with the beam, when the top-1 predicted beam is not included in beam measurements for the beam prediction.

#### FL3: RS overhead for BM-Case1

**Proposal 2-2-1a:**

* For the evaluation of the overhead for **BM-Case1**, further study the following two metrics:
  + RS overhead reduction, FFS for potential down selection:
    - Option 1:
      * where N is the number of beams (pairs) (with reference signal (SSB and/or CSI-RS)) required for measurement (in Set B)
      * where M is the total number of beams (pairs) to be predicted (in Set A)
    - Option 2:
      * where N is the number of beams (pairs) (with reference signal (SSB and/or CSI-RS)) required for measurement (in Set B)
      * where M is the total number of beams (pairs) to be predicted (in Set A)
      * FFS:
        + K is the number of Top-K selected beams (pairs) for P2 beam sweeping (if applicable)
        + K is the number of Top-K selected beams (pairs) not in Set B for P2 beam sweeping (if applicable)
    - Other options can be reported by companies
  + RS overhead, FFS for potential down selection:
    - Option 1: RS OH = N,
      * where N is the number of beams (pairs) (with reference signal (SSB and/or CSI-RS)) required for measurement (in Set B)
    - Option 2: RS OH = N + K
      * where N is the number of beams (pairs) (with reference signal (SSB and/or CSI-RS)) required for measurement (in Set B)
      * FFS:
        + K is the number of Top-K selected beams (pairs) for P2 beam sweeping (if applicable)
        + K is the number of Top-K selected beams (pairs) not in Set B for P2 beam sweeping (if applicable)
    - Other options can be reported by companies

|  |  |
| --- | --- |
| Supporting companies | NTT DOCOMO |
| Objecting companies |  |

**Proposal 2-2-2a:**

* For the evaluation of the overhead for **BM-Case2**, further study the following two metrics:
  + RS overhead reduction, FFS for potential down selection:
    - Option 1:
      * where N is the number of beams (pairs) (with reference signal (SSB and/or CSI-RS)) required for measurement (in Set B) in each slot of T1
      * where M is the total number of beams (pairs) to be predicted (in Set A) in each slot of both T1 and T2
    - Option 2:
      * where N is the number of beams (pairs) (with reference signal (SSB and/or CSI-RS)) required for measurement (in Set B) in each slot of T1
      * where M is the total number of beams (pairs) to be predicted (in Set A) in each slot of both T1 and T2
      * FFS:
        + K is the number of Top-K selected beams (pairs) for P2 beam sweeping (if applicable) in each slot of T2
        + K is the number of Top-K selected beams (pairs) not in Set B for P2 beam sweeping (if applicable) in each slot of T2
    - Other options can be reported by companies
  + RS overhead, FFS for potential down selection:
    - Option 1: RS OH = ,
      * where N is the number of beams (pairs) (with reference signal (SSB and/or CSI-RS)) required for measurement (in Set B) in each slot of T1
    - Option 2: RS OH =
      * where N is the number of beams (pairs) (with reference signal (SSB and/or CSI-RS)) required for measurement (in Set B) in each slot of T1
      * FFS:
        + K is the number of Top-K selected beams (pairs) for P2 beam sweeping (if applicable) in each slot of T2
        + K is the number of Top-K selected beams (pairs) not in Set B for P2 beam sweeping (if applicable) in each slot of T2
    - Other options can be reported by companies

|  |  |
| --- | --- |
| Supporting companies |  |
| Objecting companies |  |

**Please provide your view Proposal 2-2-1a and Proposal 2-2-2a,**

|  |  |  |
| --- | --- | --- |
| Company | Comments | |
| FL0: | FL encourages to discuss/think on the following questions:   * For RS overhead reduction in option 2, whether M includes beam sweeping, e.g., P1 +P2, especially when Set B does not belong to Set A, e.g., Set B is wide beam? * Whether the above equations can apply to both case when Set B is subset of Set A and when Set B is different from Set A? * Whether there is a need to separate the equations for DL Tx beam prediction and Tx-Rx pair prediction? * FL1: (new) M/N/K are defined with # of symbols or a number of certain type of RS? | |
| Google | We suggest we define the N/M/K based on number of symbols instead of number of beams. This could provide a common metric for all options – Tx beam prediction, Rx beam prediction and Tx-Rx beam pair prediction. In one symbol, it can be assumed that UE can try X Rx beams based on one NW Tx beam, where X is the number of UE panels. | |
| Xiaomi | For both cases, prefer to use RS overhead reduction and remove RS overhead.  For proposal 2-2-2a, both Option 1 and Option 2 can be used for the case that the periodicity of history measurement instance is same as future time instance. We suggest to add the other option for the case that the periodicity of history measurement instance is L times of the future time instance.  We suggest the following update  **Proposal 2-2-2a:**   * For the evaluation of the overhead for **BM-Case2**, further study the following two metrics:   + RS overhead reduction, FFS for potential down selection:     - The periodicity is same for measurement instance and future time instance       * Option 1-1:         + where N is the number of beams (pairs) (with reference signal (SSB and/or CSI-RS)) required for measurement (in Set B) in each ~~slot~~ measurement instance of T1         + where M is the total number of beams (pairs) to be predicted (in Set A) in each ~~slot~~ time instance of both T1 and T2       * Option 1-2:         + where N is the number of beams (pairs) (with reference signal (SSB and/or CSI-RS)) required for measurement (in Set B) in each ~~slot~~ measurement instance of T1         + where M is the total number of beams (pairs) to be predicted (in Set A) in each ~~slot~~ time instance of both T1 and T2         + FFS:   K is the number of Top-K selected beams (pairs) for P2 beam sweeping (if applicable) in each ~~slot~~ time instance of T2  K is the number of Top-K selected beams (pairs) not in Set B for P2 beam sweeping (if applicable) in each ~~slot~~ time instance of T2   * + - The periodicity of measurement instance is L (L>1) times of future time instance       * Option 2:         + Where , and L > 1.     - Other options can be reported by companies   FL1: I cannot follow your new option 2. Why this is related to RS overhead?  For changing “slot” to “time instance”, will be considered in next updates. | |
| Spreadtrum | We suggest further clarification of the definitions of T1 and T2, which may lead to ambiguity in the calculation  FL1: in the agreement in RAN 1#109e, we have the following definition for T1 and T2.   * + Where T2 is the time duration for the best beam selection, and T1 is a time duration to obtain the measurements of all the RS resource from Set B of beams. | |
| vivo | For the definition of K, our understanding is it includes two aspects   1. gNB will sweep another round of Tx beams based on the Top-K beams derived from AI model. 2. For DL Tx beam prediction based on a best Rx beam, K extra resources are needed to get the best Rx beam.   We think Aspect 2) is important to ensure fair comparison between Tx-Rx beam pair prediction and DL Tx beam prediction. Hence we suggest the following change on the definition of K for both Case 1 and Case 2   * FFS:   + K is the number of Top-K selected beams (pairs) for P2 beam sweeping (if applicable)   + K is the number of Top-K selected beams (pairs) not in Set B for P2 beam sweeping (if applicable)   + K is the number of beams used for P3 beam sweeping to get the best Rx beam (if applicable) * FFS:   + K is the number of Top-K selected beams (pairs) for P2 beam sweeping (if applicable) in each slot of T2   + K is the number of Top-K selected beams (pairs) not in Set B for P2 beam sweeping (if applicable) in each slot of T2   + K is the number of beams used for P3 beam sweeping to get the best Rx beam (if applicable)   FL1: suggestion will be considered in next update | |
| Samsung | We prefer remove RS overhead. | |
| FL1 | Please continue the discussion. | |
| HW/HiSi | We think overhead reduction in percent and overhead should both be reported. Reporting the overhead is very useful, when schemes with different size of Set A shall be compared.  Then, whether to use N or N+K, we prefer the latter. The reason is that the P2 sweeping also is a contributor to the total overhead, especially for larger numbers of K. Since K is flexible, it is also clearer to include the choice of K, when performances are compared between different schemes that use a different value of K. | |
| Futurewei | We support reporting both overhead reduction and overhead reduction % and overhead in Proposal 2-2-1a and we prefer Option 2. | |
| Qualcomm | Suggest removing RS overhead and only consider RS overhead reduction. | |
| Xiaomi | @FL, thanks for your response. Refer to R1-2209279, we consider the use case that the periodicity of history measurement instance is L times of future time instance, which can be seen in the figure below. If the periodicity of history measurement instance is 80ms, the periodicity of future time instance can be 40ms. If only one Rx beam is used for measurement, it means the periodicity for BM RS is 80ms. But AI model can predict the top-K beam with periodicity of 40ms. If without AI model, the RS need to be transmitted with 40ms period, So the RS overhead can be reduced to a half by AI model.  So we propose the Option 2.   * + - The periodicity of measurement instance is L (L>1) times of future time instance       * Option 2:         + Where , and L > 1. | |
| CAICT | RS overhead reduction is preferred. | |
| Samsung | Regarding definition of M/N/K , we prefer to keep original wording that is in terms of the number of beams. Also, we suggest to remove “K is the number of Top-K selected beams (pairs) not in Set B for P2 beam sweeping (if applicable)” in both BM-Case1 and BM-Case2 since Top-K beams can be in Set B for P2 beam sweeping. | |
| Nokia | Option 1 is needed. Option 2 only if P2 beam sweeping (two stage beam sweeping) is considered. Using only option 1 would not capture the additional overhead required for measuring the K Tx beams in case P2 beam sweeping (two stage beam sweeping) is considered.  Also, N,M,K should refer to measurements rather than beams since multiple measurements may be required for each Tx beam for the beam pair prediction. | |
| LG | Prefer to remove RS overhead. | |
| CMCC | We think RS overhead reduction is more useful since it provides the comparison with the baseline. | |
| Fujitsu | For calculation of RS overhead reduction, we need firstly get common understanding on whether RS overhead includes all P1/P2/P3 procedures of beam sweeping or only part procedure (e.g. P1 or P1/P2) of beam sweeping. After that the details of how to calculate the RS overhead reduction is discussed. | |
| ZTE | We prefer option 1. A second stage beam sweeping over the predicted top-K beams is up to gNB and should not be mandatory. Otherwise, the associated RS overhead in option 2 may be K\*U, where U is the number of UEs per cell. Besides, if set B is variable, the RS overhead reduction at the gNB side is also hard to be described by a certain equation. Regarding the definition of K, we prefer a more generic description that K is the number of beams (pairs) for obtaining a final optimal beam or beam pair based on AI model output. Besides, there is no need to separate the equations for DL Tx beam prediction and Tx-Rx pair prediction since set A/B can include Tx beam or Tx-Rx beam pair. | |
| FL2 | Please continue the discussion for BM-Case1 here for **Proposal 2-2-1b:**  **Proposal 2-2-1b:**   * For the evaluation of the overhead for **BM-Case1**, further study the following two metrics:   + RS overhead reduction, FFS for potential down selection:     - Option 1:       * where N is the number of beams (pairs) (with reference signal (SSB and/or CSI-RS)) required for measurement (in Set B)       * where M is the total number of beams (pairs) to be predicted (in Set A)     - Option 2:       * where N is the number of beams (pairs) (with reference signal (SSB and/or CSI-RS)) required for measurement (in Set B)       * where M is the total number of beams (pairs) to be predicted (in Set A)       * FFS:         + Alt1: K is the number of Top-K selected beams (pairs) for P2 beam sweeping (if applicable)         + Alt2: K is the number of Top-K selected beams (pairs) not in Set B for P2 beam sweeping (if applicable)         + Alt3: K is the number of beams used for P3 beam sweeping to get the best Rx beam (if applicable)     - Other options can be reported by companies   + RS overhead, FFS for potential down selection:     - Option 1: RS OH = N,       * where N is the number of beams (pairs) (with reference signal (SSB and/or CSI-RS)) required for measurement (in Set B)     - Option 2: RS OH = N + K       * where N is the number of beams (pairs) (with reference signal (SSB and/or CSI-RS)) required for measurement (in Set B)       * FFS:         + Alt1: K is the number of Top-K selected beams (pairs) for P2 beam sweeping (if applicable)         + Alt2: K is the number of Top-K selected beams (pairs) not in Set B for P2 beam sweeping (if applicable)         + Alt3: K is the number of beams used for P3 beam sweeping to get the best Rx beam (if applicable)     - Other options can be reported by companies | |
| Company | The support on  Opt1/Opt 2;  Alt1/Alt2/Alt3 | Comments |
| Lenovo | Prefer Option 1. | * We are OK with reporting both overhead and overhead reduction. * In proposal 2-2-1b, we prefer Option 1 for defining overhead and overhead reduction. In Option 1,   + “N” should include the total number of number of beams (with RS/SSB) that were measured for beam prediction and hence it would represent the overhead, or overhead reduction depending on the exact equation to be used. Further, when N accounts for all the measurements for predicting Top-K beams (for K=1 and K>1), it alone can represent the overhead and there is no need to further consider K.   + “M” should include the total number of available beams. In case of hierarchical beam design including wide and narrow beams, “M” should be equal to the total number of narrow beams. * Measuring N in terms of the actual time-frequency resources used for measurements also results correctly computing the overhead. We are open to further discussion on alternative definitions (such as measuring them in terms of symbols) for N, M and K.   FL 3: I think M should includes all beams not only narrow beams in option 1 |
| CATT | Option 1 is a baseline | We support to report both overhead and overhead reduction. And prefer Option 1 as baseline for defining overhead and overhead reduction in proposal 2-2-1b.  For Option2, we think it related with whether we need to perform additional beam sweeping after model inference. For example, if the AI/ML model can predict top-1 beam, which performance can be acceptable, then gNB can use this predict top-1 beam. We don’t need to perform additional beam sweeping.  Thus, to decide whether Option2 is supported, we should firstly decide whether the additional beam sweeping after model inference is necessary or not and how the additional beam sweeping is performed, e.g., P2, P3 or others. |
| MediaTek | Keep both Opt1 and Opt2 | We prefer to keep RS overhead reduction and remove RS overhead. Both Option1 and Option 2 can be kept, Option1 can be used for the case when P2 sweeping is not considered. For Alt1/2/3, agree with Fujitsu that we need to first clarify whether the overhead includes means the RS overhead for all P1/P2/P3 procedures. Otherwise, we prefer to keep Alt1/2/3 options at this stage. |
| Futurewei | Option 1 or Option 2 is ok | We support to report both the overhead and overhead reduction. We prefer Option 2 in overhead reduction and the predicted Top-K beams may incur additional sweeping overhead and should be considered. Regarding Alt1 – Alt3, we think this may be implementation dependent, we prefer leaving this to companies to describe / specify in their evaluation results. |
| Intel | Option 1 | For OH reduction, Option 1 is a baseline metric. We can consider the definition of N, M in terms of symbols to account for true OH of measurement and reporting, but current definition also works.  For Option 2, we think only Alt-2 makes sense but it may not be an accurate metric unless the UE actually performs additional measurements on these K beams. In general, the UE might use the output of the ML model to sort and select the best beam as well, in which case, K should be 0. |
| CAICT | Option 1 is preferred | At least option 1 could be used as baseline and whether other options is used could be open to discuss till the detail description of each sub use cases is clear. |
| Apple | Kept Option 1 and Option 2 | They are for different AI enabled BM schemes. In some cases, the confirmation through P2 sweeping is not needed (e.g., the beam prediction is highly accurate), in other cases, there may be such as need as confirmation through P2 sweeping. Alt. 1-3 are useful in clarifying companies’ proposals on P2 sweeping, they can be also kept. |
| vivo | Prefer Opt 2  All Alt1/Alt 2/Alt 3 can be considered | We think all the three Alts can be considered for different cases.   * Alt 1 or Alt 2 can be considered if the goal of the evaluation is to compare AI and legacy approaches. * Alt 3 is used if the goal of the evaluation is to compare Tx-Rx beam prediction and Tx beam prediction with the best Rx beam, where the resources used to get the best Rx beam need to be counted. |
| ZTE | Prefer Option 1 | Option 1 is simple to be counted and reported by companies. In option 2, RS overhead reduction should be further studied with considering following factors: the number of UEs, the beam pattern, and the refined beam sweeping procedure. For example, the transmission of N beams is cell-specific while the transmission of K beams is UE-specific. If a second stage beam sweeping over the top-K predicted beams is considered, the resulting RS overhead should be N+K\*U (instead of N+K), where U is the number of UEs per cell. Per our understanding, the second stage beam sweeping over the predicted top-K beams should be up to gNB and should not be mandatory. |
| Samsung | Keep both Option 1 and Option 2,  For Option 2, prefer Alt 1 | We’d like to prefer to focus on discussing RS overhead reduction. In SID, it clearly said “Beam management, e.g., beam prediction in time, and/or spatial domain for overhead and latency reduction, beam selection accuracy improvement [RAN1].” Referring the SID, we prefer to remove RS overhead in this proposal where we think RS overhead can be derived from RS overhead reduction by a company’s preference. Therefore, we suggest to discuss about RS overhead in the other venue.  Between Option 1 and Option 2, we think both of them can be kept for now. They may depend on the assumption on BM procedure.  For Option 2, we prefer Alt 1 since we think it can be a generic form. With Alt 1, when beam seeping after inference is not applicable, K=0, so we think it can cover Option 1 as well. For Alt 2 and Alt3, we think they do not give a generic form like Alt 1. For example, Alt 2 only can be applied when Set B is different from Set A. For the clarification, we suggest following modification:   * + - Option 2:       * where N is the number of beams (pairs) (with reference signal (SSB and/or CSI-RS)) required for measurement (in Set B)       * where M is the total number of beams (pairs) to be predicted (in Set A)       * FFS:         + Alt1: K is the number of Top-K selected beams (pairs) for P2 beam sweeping (if applicable). Otherwise, K is zero. |
| Xiaomi | Option 1 | Prefer Option 1 in OH reduction. For Option 2, if the AI model output is the RSRP of all beam pairs, additional beam sweeping of unnecessary. |
| Ericsson | Prefer Option 2, ok to keep option 1 | We think it is important to also include N+K in the KPI, otherwise there is a risk of having too optimistic results. We don’t see the difference between Alt 1 & 2 in option 2. This would be an implementation issue, and we don’t see why one would again measure beams in set B. |
| Qualcomm | OK with Option 1 and prefer Option 2 | Option 1 is primarily related to the case in which we predict the Top-1 beam (one-shot) by AI/ML inference. The evaluation results reported across the companies so far show that the Top-1 beam prediction accuracy is not quite good for beam prediction use cases (in most cases). On the other hand, Option 2 is more related to a two-step process in which Top-K beams are predicted, and then a beam sweep over the Top-K beams determines the best beam. The beam prediction accuracy of Top-K beams has been illustrated to be much more reliable compared to Top-1 across company evaluations. With this being said, we prefer Option 2. Regarding Alt 1/Alt 2/Alt 3, we do not believe the P2/P3 procedure explicitly needs to be mentioned in the alternatives and suggest rewording Alt. 2 to:  Alt2: K is the number of Top-K selected beams (pairs) not in Set B for ~~P2~~ beam sweeping (if applicable)  Prefer to keep RS overhead reduction and remove RS overhead. |
| Fujitsu | Keep both Opt1 and Opt2 | If there is no clarification on whether RS overhead includes all P1/P2/P3 procedures of beam sweeping or only part procedure (e.g. P1 or P1/P2) of beam sweeping, the both option 1 and option 2 should be kept in current stage.  And for option 2, the wording for K and Top-K may cause some confusion. It’s suggested to modify the alt1-3 like   * Alt1: K’ is the number of Top-K selected beams (pairs) for P2 beam sweeping (if applicable) * Alt2: K’ is the number of Top-K selected beams (pairs) not in Set B for P2 beam sweeping (if applicable) * Alt3: K’ is the number of beams used for P3 beam sweeping to get the best Rx beam (if applicable) |
| LG | Prefer Option 1. | It is preferred to keep RS overhead reduction and remove RS overhead. And, we think option 1 is simple and it can be considered as a baseline. |
| NTT DOCOMO | Option 2, Alt. 1/2/3 | There is a possibility that the additional measurement of the beams outside Set B is necessary for the PDSCH transmission after the beam prediction by AI/ML model. Therefore, we prefer to keep Alt.1/2/3 at this point in both RS overhead reduction and RS overhead. |
| Spreadtrum | Prefer Option 1 | We support to report both the overhead and overhead reduction. We prefer Option 1. In option 2, if k=1, RS overhead (reduction) will be calculated as N+1(1-(N+1)/M). Although 1 is not measured, it is still included in overhead, which is obviously not reasonable. I think we can make some updates to option1 to make N further include all measurement overhead.  **Proposal 2-2-1b:**   * For the evaluation of the overhead for **BM-Case1**, further study the following two metrics:   + RS overhead reduction, FFS for potential down selection:     - Option 1:       * where N is the number of beams (pairs) (with reference signal (SSB and/or CSI-RS)) required for measurement ~~(in Set B)~~       * where M is the total number of beams (pairs) to be predicted (in Set A) |
| HW/HiSi | Prefer option 2, but option 1 can be kept additionally | The overhead of P2 sweeping is important to take into consideration, especially because the K can be different for different schemes.  If no P2 sweeping is performed, then Option 1 can be used, but if 2nd round sweeping is performed, option 2 should be used:  Regarding the FFS in overhead reduction, for Option 2: we think that K is the number of Top-K selected beams (pairs) for P2 beam sweeping (Alt1)    Based on the above reasoning, we suggest the following update:  **Updated Proposal 2-2-1b:**   * For the evaluation of the overhead for **BM-Case1**, further study the following two metrics:   + When top-1 beam is inferred, RS overhead reduction, ~~FFS for potential down selection~~:     - Option 1:       * where N is the number of beams (pairs) (with reference signal (SSB and/or CSI-RS)) required for measurement (in Set B)       * where M is the total number of beams (pairs) to be predicted (in Set A)     - when top-K, K>1, beams are inferred, Option 2:       * where N is the number of beams (pairs) (with reference signal (SSB and/or CSI-RS)) required for measurement (in Set B)       * where M is the total number of beams (pairs) to be predicted (in Set A)       * FFS:         + Alt1: K is the number of Top-K selected beams (pairs) for P2 beam sweeping (if applicable)         + ~~Alt2: K is the number of Top-K selected beams (pairs) not in Set B for P2 beam sweeping (if applicable)~~         + ~~Alt3: K is the number of beams used for P3 beam sweeping to get the best Rx beam (if applicable)~~     - Other options can be reported by companies   + RS overhead, ~~FFS for potential down selection:~~     - When top-1 beam is inferred Option 1: RS OH = N,       * where N is the number of beams (pairs) (with reference signal (SSB and/or CSI-RS)) required for measurement (in Set B)     - when top-K, K>1, beams are inferred ,Option 2: RS OH = N + K       * where N is the number of beams (pairs) (with reference signal (SSB and/or CSI-RS)) required for measurement (in Set B)       * FFS:         + Alt1: K is the number of Top-K selected beams (pairs) for P2 beam sweeping (if applicable)         + ~~Alt2: K is the number of Top-K selected beams (pairs) not in Set B for P2 beam sweeping (if applicable)~~         + ~~Alt3: K is the number of beams used for P3 beam sweeping to get the best Rx beam (if applicable)~~   Other options can be reported by companies |
| CMCC | Option 2, Alt. 1/2/3 | In our view, Alt1 and Alt2 can be merged into one option as follows:   * + - * + Alt1/Alt2: K is the number of Top-K selected beams (pairs) not in Set B for ~~P2~~ beam sweeping (if applicable) |
| FL3 |  | Thanks for the good study. I think for this meeting, we can keep both options and RS overhead reduction and RS overhead, as well as three alternatives. In my understanding, they may be useful in different cases in companies’ mind.  Please check the following proposal with changes marked in red.  **Proposal 2-2-1c:**   * For the evaluation of the overhead for **BM-Case1**, further study the following two metrics:   + RS overhead reduction, FFS for potential down selection:     - Option 1:       * where N is the number of beams (pairs) (with reference signal (SSB and/or CSI-RS)) required for measurement (in Set B)       * where M is the total number of beams (pairs) to be predicted (in Set A)     - Option 2:       * where N is the number of beams (pairs) (with reference signal (SSB and/or CSI-RS)) required for measurement (in Set B)       * where M is the total number of beams (pairs) to be predicted (in Set A)       * FFS:         + Alt1: P is the number of Top-K selected beams (pairs) for ~~P2~~ beam sweeping (if applicable)         + Alt2: P is the number of Top-K selected beams (pairs) not in Set B for ~~P2~~ beam sweeping (if applicable)         + Alt3: P is the number of beams used for ~~P3~~ beam sweeping to get the best Rx beam (if applicable)     - Other options can be reported by companies   + RS overhead, FFS for potential down selection:     - Option 1: RS OH = N,       * where N is the number of beams (pairs) (with reference signal (SSB and/or CSI-RS)) required for measurement (in Set B)     - Option 2: RS OH = N + P       * where N is the number of beams (pairs) (with reference signal (SSB and/or CSI-RS)) required for measurement (in Set B)       * FFS:         + Alt1: P is the number of Top-K selected beams (pairs) for ~~P2~~ beam sweeping (if applicable)         + Alt2: P is the number of Top-K selected beams (pairs) not in Set B for ~~P2~~ beam sweeping (if applicable)         + Alt3: P is the number of beams used for ~~P3~~ beam sweeping to get the best Rx beam (if applicable)     - Other options can be reported by companies |
| InterDigital |  | **Overhead reduction**  We are fine to keep both option 1 and option 2.  However, we have some questions on option 2.  Why do we need additional beam sweeping with P? If we remove P2, then the motivation of supporting additional beam sweeping with P is not clear enough to us.  In addition, the denominator of option 2 should be changed if we add P to the numerator. For example, if additional beam sweeping for P3 is considered for the spatial domain beam prediction, additional beam sweeping for P3 should be considered for the baseline as well.  **RS Overhead**  For us, this metric is not clear enough. For overhead reduction, we can only consider number of beams by using the assumption that the numerator and the denominator use same RS overhead. However, if we want to count the actual RS overhead, the overhead should be calculated based on the actual RS overhead not using the number of beams. |
| Xiaomi |  | Support the proposal 2-2-1c and prefer to remove RS overhead. |
| OPPO |  | It seems the RS overhead is reflected in the RS overhead reduction. To avoid redundancy, keeping only the RS overhead reduction seems okay.  As for those alternatives of P, in our view, how to calculate P also depends on certain prerequisite, such as Tx beam or Tx-Rx beam pair prediction, and the strategy to find/confirm beams in the 2nd round of beam sweeping. It can be complicated and case-by-case. We are fine to study and hopefully some pre-conditions can be listed along with each alternative. |
| HW/HiSi |  | We are fine to keep both options on the table, and we still think that overhead also should be reported for better comparison across schemes with different size of set A. |
| CATT |  | Fine with the proposal 2-2-1c. Delete the RS overhead part is also OK for us. |
| Samsung |  | We are fine with keeping both RS overhead reduction and RS overhead at least at this meeting. However, since some companies may not prefer to discuss RS overhead together, we think it would be better to discuss RS overhead in the other proposal. Therefore, we suggest to update the proposal as follows:  **Proposal 2-2-1~~c~~-1a:**   * For the evaluation of the overhead for **BM-Case1**, further study ~~the following two metrics:~~RS overhead reduction, FFS for potential down selection:   + Option 1:     - where N is the number of beams (pairs) (with reference signal (SSB and/or CSI-RS)) required for measurement (in Set B)     - where M is the total number of beams (pairs) to be predicted (in Set A)   + Option 2:     - where N is the number of beams (pairs) (with reference signal (SSB and/or CSI-RS)) required for measurement (in Set B)     - where M is the total number of beams (pairs) to be predicted (in Set A)     - FFS:       * Alt1: P is the number of Top-K selected beams (pairs) for ~~P2~~ beam sweeping (if applicable)       * Alt2: P is the number of Top-K selected beams (pairs) not in Set B for ~~P2~~ beam sweeping (if applicable)       * Alt3: P is the number of beams used for ~~P3~~ beam sweeping to get the best Rx beam (if applicable)     - Other options can be reported by companies   **Proposal 2-2-1-2a:**   * For the evaluation of the overhead for **BM-Case1**, further study the necessity of RS overhead, and FFS for potential down selection:   + Option 1: RS OH = N,     - where N is the number of beams (pairs) (with reference signal (SSB and/or CSI-RS)) required for measurement (in Set B)   + Option 2: RS OH = N + P     - where N is the number of beams (pairs) (with reference signal (SSB and/or CSI-RS)) required for measurement (in Set B)     - FFS:       * Alt1: P is the number of Top-K selected beams (pairs) for ~~P2~~ beam sweeping (if applicable)       * Alt2: P is the number of Top-K selected beams (pairs) not in Set B for ~~P2~~ beam sweeping (if applicable)       * Alt3: P is the number of beams used for ~~P3~~ beam sweeping to get the best Rx beam (if applicable)   + Other options can be reported by companies   Regarding Alt 2 in Option 2, we think Alt 2 is subset of Alt 1 and Alt 1 can be considered as a lower bound, since we cannot guarantee measured beams in Set B will not be blocked in the next measurement time instance. |
| LG |  | Support in principle, but, we also prefer to remove RS overhead. |
| NTT DOCOMO |  | We are fine to keep both Option 1 and Option 2. |
| Spreadtrum |  | Fine with the proposal 2-2-1c. |
| Lenovo |  | We are fine with proposal 2-2-1c. |
| Ericsson |  | Support the proposal 2-2-1c, ok to remove RS overhead if it is the majority view |
| MediaTek |  | Support Proposal 2-2-1c and we prefer to remove RS overhead. Also, we would like to confirm regarding Alt.3. Is Alt.3 only applicable to the evaluation of Tx/Rx beam pair prediction and Rx beam prediction, not applicable to the evaluation for Tx beam prediction? Regarding the 3 alternatives, to avoid confusion, we agree with OPPO that it is better to discuss the usage of each alternative in separate the cases (e.g. Tx beam prediction, beam pair prediction). For example, for Tx beam prediction evaluation, Alt1 and Alt2 can be used but Alt3 seems to be not applicable. |

#### FL3: RS overhead for BM-Case2

**Proposal 2-2-2b:**

* For the evaluation of the overhead for **BM-Case2**, further study the following two metrics:
  + RS overhead reduction, FFS for potential down selection:
    - Option 1-1:
      * where N is the number of beams (pairs) (with reference signal (SSB and/or CSI-RS)) required for measurement (in Set B) in each slot of T1
      * where M is the total number of beams (pairs) to be predicted (in Set A) in each slot of both T1 and T2
    - Option 1-2:
      * where N is the number of beams (pairs) (with reference signal (SSB and/or CSI-RS)) required for measurement (in Set B) in each slot of T1
      * where M is the total number of beams (pairs) to be predicted (in Set A) in each slot of both T1 and T2
      * FFS:
        + Alt1: K is the number of Top-K selected beams (pairs) for P2 beam sweeping (if applicable) in each slot of T2
        + Alt2: K is the number of Top-K selected beams (pairs) not in Set B for P2 beam sweeping (if applicable) in each slot of T2
        + Alt 3: K is the number of beams used for P3 beam sweeping to get the best Rx beam (if applicable)
    - Option 2:
      * where N is the number of beams (pairs) (with reference signal (SSB and/or CSI-RS)) required by scheme with AI
      * where M is the number of beams (pairs) (with reference signal (SSB and/or CSI-RS)) required by baseline scheme
    - Other options can be reported by companies
  + RS overhead, FFS for potential down selection:
    - Option 1: RS OH = ,
      * where N is the number of beams (pairs) (with reference signal (SSB and/or CSI-RS)) required for measurement (in Set B) in each slot of T1
    - Option 2: RS OH =
      * where N is the number of beams (pairs) (with reference signal (SSB and/or CSI-RS)) required for measurement (in Set B) in each slot of T1
      * FFS:
        + Alt1: K is the number of Top-K selected beams (pairs) for P2 beam sweeping (if applicable) in each slot of T2
        + Alt2: K is the number of Top-K selected beams (pairs) not in Set B for P2 beam sweeping (if applicable) in each slot of T2
        + Alt 3: K is the number of beams used for P3 beam sweeping to get the best Rx beam (if applicable)
    - Other options can be reported by companies

**Please provide your view Proposal 2-2-2b:**

|  |  |  |
| --- | --- | --- |
| Company | Comments | |
| FL2: | Please continue the discussion for BM-Case2 here for **Proposal 2-2-2b:** | |
| Company | The support on  Opt1/Opt 2;  Alt1/Alt2/Alt3 | Comments |
| Lenovo | Option 1-1 |  |
| CATT |  | For Option 1-1, why there is T2 in SUM of M in the denominator. We think T2 is related with the additional beam sweeping after the model inference. But for Option1-1, there is no additional beam sweeping. Is that correct understanding? If so, we propose to change Option 1-1 as following:   * + - Option 1-1:       * where N is the number of beams (pairs) (with reference signal (SSB and/or CSI-RS)) required for measurement (in Set B) in each slot of T1       * where M is the total number of beams (pairs) to be predicted (in Set A) in each slot of ~~both~~ T1 ~~and T2~~   We support the above updated Option1-1 for RS overhead reduction and Option 1 for RS overhead as a baseline. For Option1-2 for RS overhead reduction and Option 2 for RS overhead, we have same concern as Proposal 2-2-1b. We should firstly decide whether the additional beam sweeping after model inference is necessary or not and how the additional beam sweeping is performed, e.g., P2, P3 or others. |
| MediaTek | Keep both Opt1 and Opt2 | We prefer to keep RS overhead reduction and remove RS overhead. For RS overhead reduction, both Option1 and Option 2 can be kept. |
| vivo | Prefer Opt 2  All Alt1/Alt 2/Alt 3 can be considered | We think all the three Alts can be considered for different cases.   * Alt 1 or Alt 2 can be considered if the goal of the evaluation is to compare AI and legacy approaches. * Alt 3 is used if the goal of the evaluation is to compare Tx-Rx beam prediction and Tx beam prediction with the best Rx beam, where the resources used to get the best Rx beam need to be counted. |
| Samsung | Keep both options | We suggest to focus the discussion for BM-Case1 first, and then to extend it to BM Case 2. |
| Xiaomi |  | We prefer OH reduction.  And for the case that the periodicity of history measurement instance is same as future time instance, we prefer Option 1-1.  While for the case that the periodicity of history measurement instance is L times of the future time instance. We prefer the following definition.   * + - The periodicity of measurement instance is L (L>1) times of future time instance       * Option 2:         + Where , and L > 1.   Other options can be reported by companies |
| Ericsson | Prefer Option 2, ok to keep option 1 | Share the view by Samsung |
| Qualcomm | OK with Option 1-1 and prefer Option 1-2 | Option 1-1 is primarily related to the case in which we predict the Top-1 beam (one-shot) by AI/ML inference. The evaluation results reported across the companies so far show that the Top-1 beam prediction accuracy is not quite good for beam prediction use cases (in most cases). On the other hand, Option 1-2 is more related to a two-step process in which Top-K beams are predicted, and then a beam sweep over the Top-K beams determines the best beam. The beam prediction accuracy of Top-K beams has been illustrated to be much more reliable compared to Top-1 across company evaluations. With this being said, we prefer Option 1-2. Regarding Alt 1/Alt 2/Alt 3, we do not believe the P2/P3 procedure explicitly needs to be mentioned in the alternatives and suggest rewording Alt. 2 to:   * + - * + Alt2: K is the number of Top-K selected beams (pairs) not in Set B for ~~P2~~ beam sweeping (if applicable) in each slot of T2   Prefer to keep RS overhead reduction and remove RS overhead. |
| LG | Keep both option2 | It is preferred to keep RS overhead reduction and remove RS overhead. |
| NTT DOCOMO | Option 1-2 with Alt.1/2/3;  Option 2 | We support both Option 1 and 2 to cover the same/different periodicity of measurements and predictions. Within Option 1, our preference is Option 1-2 with Alt.1/2/3 due to the same reason as Proposal 2-2-1b. |
| HW/HiSI |  | Option 2 and Option 1, similar to our reasoning for BM-Case 1.  But we suggest that we should postpones the discussion of BM-Case 2, until we have agreed BM-Case 1. It will become much simpler then, since BM-Case 2 is an extension of Case 1. |
| CMCC | Option 1-2 | Similar comments as Proposal 2-2-1b. Alt1 and Alt2 can be merged into one Alt. |
| FL3 |  | Assuming Proposals for BM-Case1 is table, companies are encouraged to provide comments for BM-Case2.  In my understanding, Xiaomi’s proposal for RS overhead reduction make sense. However, I feel we might need more time to study offline. In this meeting, the intention is list options on the table, then we can choose a proper definition for the result collection in next meeting.  **Proposal 2-2-2c:**   * For the evaluation of the overhead for **BM-Case2**, further study the following two metrics:   + RS overhead reduction, FFS for potential down selection:     - Option 1-1:       * where N is the number of beams (pairs) (with reference signal (SSB and/or CSI-RS)) required for measurement (in Set B) in each slot of T1       * where M is the total number of beams (pairs) to be predicted (in Set A) in each slot of both T1 and T2     - Option 1-2:       * where N is the number of beams (pairs) (with reference signal (SSB and/or CSI-RS)) required for measurement (in Set B) in each slot of T1       * where M is the total number of beams (pairs) to be predicted (in Set A) in each slot of both T1 and T2       * FFS:         + Alt1: P is the number of Top-K selected beams (pairs) for ~~P2~~ beam sweeping (if applicable) in each slot of T2         + Alt2: P is the number of Top-K selected beams (pairs) not in Set B for ~~P2~~ beam sweeping (if applicable) in each slot of T2         + Alt 3: P is the number of beams used for ~~P3~~ beam sweeping to get the best Rx beam (if applicable)     - Option 2:       * where N is the number of beams (pairs) (with reference signal (SSB and/or CSI-RS)) required by scheme with AI       * where M is the number of beams (pairs) (with reference signal (SSB and/or CSI-RS)) required by baseline scheme     - Other options can be reported by companies   + RS overhead, FFS for potential down selection:     - Option 1-1: RS OH = ,       * where N is the number of beams (pairs) (with reference signal (SSB and/or CSI-RS)) required for measurement (in Set B) in each slot of T1     - Option 1-2: RS OH =       * where N is the number of beams (pairs) (with reference signal (SSB and/or CSI-RS)) required for measurement (in Set B) in each slot of T1       * FFS:         + Alt1: P is the number of Top-K selected beams (pairs) for ~~P2~~ beam sweeping (if applicable) in each slot of T2         + Alt2: P is the number of Top-K selected beams (pairs) not in Set B for ~~P2~~ beam sweeping (if applicable) in each slot of T2         + Alt 3: P is the number of beams used for ~~P3~~ beam sweeping to get the best Rx beam (if applicable)     - Other options can be reported by companies |
| InterDigital |  | We also prefer to focus on BM Case-1 first. |
| Xiaomi |  | We are OK with proposal 2-2-2c, and support both Option 1 and Option 2. Option 1-1 and Option 1-2 can be used for same periodicity for history measurement instance and future predicted time instance, and Option 2 can be used for different periodicity. |
| OPPO |  | Generally fine and similar comments as we leave for BM-Case1.  By checking Option 1-1/Option 1-2 and Option 2, it seems Option 2 can only roughly capture the RS overhead reduction whereas Option 1-x can capture the overhead in a more refined granularity, e.g. per slot. |
| HW/HiSi |  | Support IDC, BM-Case 1 should be defined firstly. This will make it easier for BM-Case 2. We can leave the options here for now and come back next meeting on BM Case-2 |
| CATT |  | We wonder why T2 is in the Option 1-1 for RS overhead reduction.  In our previous comments, we propose to change the Option1-1 as following:   * + - Option 1-1:       * where N is the number of beams (pairs) (with reference signal (SSB and/or CSI-RS)) required for measurement (in Set B) in each slot of T1       * where M is the total number of beams (pairs) to be predicted (in Set A) in each slot of ~~both~~ T1 ~~and T2~~ |
| LG |  | We also fine to focus on BM Case-1. |
| NTT DOCOMO |  | We are fine with proposal 2-2-2c. Within Option 1, we prefer Option 1-2.  Option 2 is trying to capture the case where inputs and outputs of AI/ML model are with different periodicities instead of only roughly capturing the RS overhead reduction. |
| Spreadtrum |  | Support IDC, we can discuss the options for BM Case-2 until the options for BM Case-1 are stable. |
| MediaTek |  | Similar comments as we have for BM-Case1. We prefer removing RS overhead and keeping both Options 1 and 2 in RS overhead reduction. Alt1/2/3 can be discussed case by case. |

#### FL3: QCL relation

* DoCoMo [25]:
  + Observation 1: Additional beam measurements might be necessary for PDSCH/PDCCH reception with top1/K predicted beam(s), when the top-1/K predicted beam(s) are not included in beams measured for the beam prediction.
  + Proposal 1: Discuss the requirement of actual QCL relation, and consider the additional RS measurement overhead to obtain the actual QCL relation if necessary.



Figure 1. Additional beam measurements with the top-1 predicted beam for reception with the beam, when the top-1 predicted beam is not included in beam measurements for the beam prediction.

**Please provide your view on additional RS measurement overhead to obtain the actual QCL relation?**

|  |  |
| --- | --- |
| Company | Comments |
| FL0 | This might be a valid issue. However, in FL’s views, there is no urgency to discuss addition RS for P3 beam sweeping for this meeting. However, companies are encouraged to describe the assumption of P1/P2/P3 in their simulation assumption, and provide analysis on RS overhead reduction accordingly. For example, whether all of the procedure or only part of the procedure is considered, QCL assumption, reuse some of legacy procedures, etc. |
| Google | We assume this is for QCL-TypeD measurement instead of QCL-TypeA. For QCL-TypeA, there is no additional overhead, since TRS should always be transmitted. For QCL-TypeD, it is true that additional measurement is needed if the beam is unknown. |
| vivo | Isn’t this included in the discussion of K from the previously proposal? |
| FL1 | Please continue the discussion. |
| NTT DOCOMO | Agree with Google in the point that the additional measurements might be necessary for QCL type D. If the additional measurement is necessary for QCL typeD, we should consider the additional measurements in RS measurement overhead KPI calculation. This additional measurement should be considered regardless of topK or top1 beam prediction, because the actual L1-RSRP cannot be obtained even for top1 beam prediction.  We would like to hear more views from companies. |
| Nokia | There are the extra steps of beam measurements prior to beam indication of a predicted beam (in case those predicted beams are not getting measured at least with a longer periodicity). |
| CMCC | We also think this should be part discussion of the previous proposal. We think this is the reason to include K in Option 2 of the definition of RS overhead reduction. |
| Fujitsu | In our understanding, such additional measurements to obtain QCL relation for QCL-TypeD are necessary. |
| Intel | This should be a valid issue for QCL-Type D but we think this issue would impact baseline as well in case the baseline is assumed to be hierarchical search. |
| ZTE | We agree with DCM. However, additional measurements might not always be necessary for obtaining an actual QCL-TypeD relation. For example, in the Tx beam prediction case for P2 procedure, all narrow Tx beams (including the unmeasured top-1 predicted beam) may be refined from a same wide beam, and thus can be configured with the same QCL-TypeD relation. In this case, additional measurements is not needed. |
| FL2 | In my view, I would like to encourage companies to consider using AI to reply one step of BM first, with explanation on other step(s). For example, using RS overhead (reduction) option 1, to explain how much AI can help, assuming, e.g., gNB can select the best beam with “limited” report, (e.g., in P2).  Then next step, we can see how/how much AI can help in the whole procedure, e.g., using RS overhead option 2 consider sweeping with “K. and, even consider AI can be used at multiple steps. |
| CATT | We think we need first discuss the whole procedure for BM when AI/ML involve. We think the initial motivation of study item is to use AI mechanism to instead the whole steps of traditional BM. If this is not feasible, then we can discuss whether we need additional pre- and post-processing of beam sweeping, e.g., P2 or P3. Maybe in some cases the additional beam sweeping is needed to obtain QCL relation. Then we can have some conclusion to say how much AI can help for BM. |
| InterDigital | This may be valid issue if Set A and Set B are different. However, in our view, clear usage scenarios for the case that Set A and Set B are not identified yet. Having said that, we prefer to discuss this issue after clarifying the use cases for Set A ≠ Set B. |
| Samsung | In our view, narrow beams in Set A can be QCLed with wide beams (e.g., SSB) in Set B so there may be no unknown TCI state condition issues to get L1-RSRP measurements for the RS QCLed to the target TCI state. Also, we believe whether to use of additional procedure to get QCL-type D of predicted narrow beams can be up to NW’s decision. |
| MediaTek | We also think this issue is valid, but it depends on scenarios. For certain scenarios, we may or may not have this problem. For example, when Set B is a subset of Set A, if the predicted Top-1 beam is not in Set B, then we need to have additional measurement for QCL relation. We can first study and list the scenarios when such QCL relation overhead exists, then discuss how to capture it. |

### 2.2.3 UCI report

In RAN 1 #110, the following agreement on UCI report was agreed.

|  |
| --- |
| **Agreement**   * **To evaluate the performance of AI/ML in beam management at least for NW side beam prediction, UCI report overhead can be further studied as one of KPI options.**    + **FFS: number of UCI reports and UCI payload size** |

The following was discussed in contributions:

* Vivo [5]:
  + Proposal 8: UCI reporting overhead reduction, including the number of UCI report and UCI payload size, should be considered as basic KPI.
* Lenovo [15]:
  + To account for the other kind of overhead, it is required to account for the number of UCI reports and the size of each UCI report (in bits). These quantities (i.e., the no. of UCI reports and the size of such reports) need to be compared with the case of exhaustive search for arriving at a meaningful measure of the amount of reporting overhead reduction offered by the AI/ML model under consideration.
  + Any other signals that need to be exchanged between UE and gNB to support the AI/ML model, such as signaling in another carrier (e.g., FR1), UE location information, spatial features of the environment etc., should also be considered accounted for.
* DoCoMo [25]:
  + Proposal 2: Consider the number of transmissions for UCI as performance KPI:
    - It is beneficial to reduce the number of uplink transmissions for commercial aspects
    - Temporal beam prediction with NW side model can enable beam management with low frequent beam measurement reports

#### FL2: (close)UCI report

**Proposal 2-2-3a: (updated from Agreement in RAN 1 #110)**

* **To evaluate the performance of AI/ML in beam management at least for NW side beam prediction, UCI report overhead can be further studied as one of KPI options.** 
  + **~~FFS:~~ number of UCI reports and/or UCI payload size for each prediction**

|  |  |
| --- | --- |
| Supporting companies | MediaTek, NTT DOCOMO, Lenovo, vivo, CAICT |
| Objecting companies |  |

**Please provide your view Proposal 2-2-3a:**

|  |  |
| --- | --- |
| Company | Comments |
| Google | We are ok with the proposal in principle, but it is hard to determine the UCI overhead with regard to different UCI quantization schemes. |
| LG | Agree with Google. |
| HW/HiSi | Seems ok on first sight, but no need to hurry resolving the FFS either. We would prefer to postpone this issue. |
| Spreadtrum | Agreeing with Google, the overhead of UCI is hard to quantify |
| Nokia | Ok in general. |
| NTT DOCOMO | The number of uplink transmissions for UCI is an important commercial aspect. The gain of this point should be evaluated for NW side beam prediction. |
| Lenovo | We think UCI report overhead is an important metric. Propose to revise the sub-bullet as follows:  **~~FFS:~~ number of UCI reports and~~/or~~ UCI payload size for each prediction** |
| Ericsson | Agree in general, but we don’t see why we need to agree on this at this stage. It is also unclear what the sub bullet refers to (assume it is the definition of UCI report overhead). |
| Samsung | We share the same view with HW/HiSi. |
| CATT | Agree in general. But currently we prefer to keep the FFS in sub-bullet. |
| FL1 | Please continue the discussion. |
| NVIDIA | Fine to study UCI overhead and agree with HW/HiSi that FFS can be kept (or simply removed) |
| Apple | It seems premature to settle down on details, so keeping the FFS wording is preferred. |
| OPPO | We are fine with the study in general.  But before we step down the payload of UCI, we think first for NW-side beam prediction, we may need to decide whether/how to report the whole measurements of Set B for BM-Case1 and BM-Case2. |
| CMCC | How to evaluate UCI overhead should be clarified, since different reporting method can be used. Prefer to keep the FFS. |
| Fujitsu | We are fine for UCI report overhead as one of KPI. However, how to define the number of UCI reports for each prediction is not clear. And the UCI payload size is difficult to determine in current stage since the number of measurement results and quantization bits for each result within a report may be enhanced for beam prediction on AI/ML based on legacy UCI report format. |
| ZTE | We prefer to postpone the UCI related discussion since the construction of set A/B has not been converged and whether to do further reporting compression or optimization is not clear. |
| MediaTek | Agree with HW/HiSi and Google. We can postpone this discussion until there is a consensus on UCI report format. |
| Intel | Can be considered as a KPI but it is not easy at this stage to determine UCI payload without accurate idea of specification impact. |
| HW/HiSi | We also think that UCI reporting overhead is important, but for the efficiency of discussion we still believe it is better to postpone. |
| FL3 | Let’s further discuss this in later meeting. |

### 2.2.3 Other KPIs

The following other KPIs were proposed in the contributions:

RRC signaling overhead:

* Vivo [5]:
  + Proposal 9: RRC singling overhead can be considered as optional KPI if huge amount of data, such as training data, assistant information, and AI model data, is exchanged via RAN air interference.

Latency reduction

* Interdigital [6]
  + Proposal 4: Reporting overhead and latency aspects should be considered in evaluation of system performance not as independent KPIs.
* Lenovo [15]
  + Proposal 4 Consider Beam Prediction Accuracy, Overhead Reduction and Latency Reduction as the key KPIs in evaluating an AI/ML model for beam management and consider adopting the definitions proposed above.

FL0: no urgency to discuss other KPIs in this meeting.

## 2.3 (Closed) Model size and computational complexity

Several companies proposed to consider model size and computation complexity for AI/ML model.

* Ericsson [11]
  + Proposal 2: When presenting results for AI/ML models, the proponent should report a model size (e.g., number of parameters) and an estimate of the number of floating-point operations (FLOPs) for inference.
* MTK [20]
  + Proposal 1: For AI/ML-based beam prediction evaluation, adopt the FLOPs and/or MACs as the time complexity, and the number of parameters as the space complexity, other options are not precluded.
* NVIDA [23]
  + Proposal 3: For evaluation of AI/ML based beam management, the computational complexity can be reported via the metric of floating point operations (FLOPs) for inference.
  + Proposal 4: For evaluation of AI/ML based beam management, the model complexity may be measured by memory storage in terms of number of AI/ML model parameters.
  + Observation 1: Increasing hardware performance can support successively more complex AI/ML models. For example, GPU inference performance has improved by 317x in 8 years (2012-2020), more than doubling each year.
  + Proposal 5: AI/ML model complexity and computational complexity should not be regarded as a roadblock to the adoption of AI/ML based algorithms for beam management enhancements.

FL0: In FL’s understanding, the agreements in 9.2.1 also apply to beam management. Therefore, no need to reopen the discussion here.

|  |
| --- |
| **Agreement**  *The following is an initial list of common KPIs (if applicable) for evaluating performance benefits of AI/ML*   * *Performance*   + *Intermediate KPIs*   + *Link and system level performance*   + *Generalization performance* * *Over-the-air Overhead*   + *Overhead of assistance information*   + *Overhead of data collection*   + *Overhead of model delivery/transfer*   + *Overhead of other AI/ML-related signaling* * *Inference complexity*   + *Computational complexity of model inference: FLOPs*   + *Computational complexity for pre- and post-processing*   + *Model complexity: e.g., the number of parameters and/or size (e.g. Mbyte)* * *Training complexity* * *LCM related complexity and storage overhead*   + *FFS: specific aspects* * *FFS: Latency, e.g., Inference latency*   *Note: Other aspects may be added in the future, e.g. training related KPIs*  *Note: Use-case specific KPIs may be additionally considered for the given use-case.* |

## 2.4 (Closed) Baseline performance

Some companies provided some analysis on baseline performance for benchmark.

* Huawei/HiSi [2]
  + Proposal 15: For spatial domain beam prediction, both of the two baselines for performance evaluation shall be considered:
    - An upper performance bound obtained by exhaustive sweep over Set A
    - A lower performance bound obtained by non-AI/ML-based legacy sparse beam sweeping with the same overhead as the AI/ML-based approach
* Vivo [3]
  + Proposal 3: Support both option 1 and option 2 as baseline performance in spatial domain beam prediction and temporal domain beam prediction, and set B selection method in option 2 should be reported.
* InterDigital [6]
  + Observation 1: Legacy beam management with Rel-17 without AI/ML algorithms is not an appropriate baseline as implementation-based AI/ML operation is available for UE and gNB implementations.
  + Proposal 1: ‘No collaboration framework: AI/ML algorithms purely implementation based and not requiring air-interface changes’ could be an appropriate baseline to accurately evaluate the benefits of AI/ML with specification enhancements.
  + FL0: based on current agreements, company can report the conventional scheme as baseline.
* OPPO [8]
  + Proposal 9: For spatial domain beam prediction, select the best beam within Set A via exhaustive beam sweeping (Option 1) as baseline.
  + Proposal 10: For temporal domain beam prediction, select the best beam for T2 within Set A via exhaustive beam sweeping (Option 1a) as baseline.
* Google [9]
  + Proposal 3: For spatial-domain beam prediction, the baseline performance should be the performance from the beam selected from set B beams.
  + Proposal 4: For time-domain beam prediction, the baseline performance should be the performance without beam change for T2, i.e. the beam used prior to T2 is applied for T2.
* Intel [14]
  + For baseline performance evaluation, Option 2 should correspond to hierarchical beam search where, based on sub-use case being evaluated, set B may be a subset of set A or set B can contain both wide and correlated narrow beams.

FL0: There is no strong need to down select the baseline performance in this meeting other than current agreements as below:

|  |
| --- |
| **Agreement**   * For spatial-domain beam prediction, further study the following options as baseline performance   + Option 1: Select the best beam within Set A of beams based on the measurement of all RS resources or all possible beams of beam Set A (exhaustive beam sweeping)      - FFS CSI-RS/SSB as the RS resources   + Option 2: Select the best beam within Set A of beams based on the measurement of RS resources from Set B of beams     - FFS: Set B is a subset of Set A and/or Set A consists of narrow beams and Set B consists of wide beams     - FFS: how conventional scheme to obtain performance KPIs     - FFS: how to determine the subset of RS resources is reported by companies   + Other options are not precluded.   **Agreement**   * For temporal beam prediction, further study the following options as baseline performance   + Option 1a: Select the best beam for T2 within Set A of beams based on the measurements of all the RS resources or all possible beams from Set A of beams at the time instants within T2   + Option 2: Select the best beam for T2 within Set A of beams based on the measurements of all the RS resources from Set B of beams at the time instants within T1     - Companies explain the detail on how to select the best beam for T2 from Set A based on the measurements in T1   + Where T2 is the time duration for the best beam selection, and T1 is a time duration to obtain the measurements of all the RS resource from Set B of beams.     - T1 and T2 are aligned with those for AI/ML based methods   + Whether Set A and Set B are the same or different depend on the sub-use case   + Other options are not precluded. |

# AI/ML model Generalization

Generalization is one of the important aspects to verify the performance of AI/ML model.

## 3.1 Evaluation assumption for generalization performance

The follow discussions/proposals were summarized:

**General principle**

* Futurewei [1]
  + Table 4.1-1: Model generalization evaluation report example

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Sub use case** | **Training scenario / config.** | **Testing scenario/ config.** | **Set A/B configurations** | **Dataset size** | | **Perf. KPIs** | **Other KPIs** | **Mechanism applied** |
| Train | Test |
|  |  |  |  |  |  |  |  |  |

* + Proposal 3: When reporting AI/ML model generalization evaluation results for beam management enhancements, companies are encouraged to align the reporting attributes and format as depicted in Table 4.1-1.
  + Proposal 4: In AI/ML model generalization across different scenarios/configurations for spatial-domain beam prediction sub use case, further study the applicable generalization mechanism(s) that can be applied to different scenario/configuration combinations.
* Huawei/HiSi [2]:
  + Proposal 4: To verify the generalization of AI/ML models on AI/ML-based beam management in both spatial and temporal domain, the following cases to construct the training dataset and testing dataset should be considered:
    - Case 1: The AI/ML model is trained based on training dataset from one Scenario#A/Configuration#A, and then the AI/ML model is tested on dataset from the same Scenario#A/Configuration#A
    - Case 2: The AI/ML model is trained based on training dataset from one Scenario#A/Configuration#A, and then the AI/ML model is tested on dataset from a different Scenario#B/Configuration#B
    - Case 2A: The AI/ML model is trained based on training dataset from one Scenario#A/Configuration#A, and then the AI/ML model is fine-tuned based on the fine-tuning dataset from a different Scenario#B/Configuration#B. After that, the AI/ML model is tested on dataset subject to the same Scenario#B/Configuration#B as the fine-tuning dataset
    - Case 3: The AI/ML model is trained based on training dataset constructed by mixing datasets from multiple scenarios including Scenario#A/Configuration#A and Scenario#B/Configuration#B, and then the AI/ML model is tested on dataset from a single Scenario#A/Configuration#A or Scenario#B/Configuration#B from the multiple scenarios
* Vivo [5]
  + Proposal 10: Support to define generalization performance KPI.
* China Telecom[7]
  + Proposal 3: The following cases are considered for verifying the generalization performance of an AI/ML model over various scenarios/configurations as a starting point:
    - Case 1: The AI/ML model is trained based on training dataset from one Scenario#A/Configuration#A, and then the AI/ML model performs inference/test on a dataset from the same Scenario#A/Configuration#A
    - Case 2: The AI/ML model is trained based on training dataset from one Scenario#A/Configuration#A, and then the AI/ML model performs inference/test on a different dataset than Scenario#A/Configuration#A, e.g., Scenario#B/Configuration#B, Scenario#A/Configuration#B
    - Case 3: The AI/ML model is trained based on training dataset constructed by mixing datasets from multiple scenarios/configurations including Scenario#A/Configuration#A and a different dataset than Scenario#A/Configuration#A, e.g., Scenario#B/Configuration#B, Scenario#A/Configuration#B, and then the AI/ML model performs inference/test on a dataset from a single Scenario/Configuration from the multiple scenarios/configurations, e.g., Scenario#A/Configuration#A, Scenario#B/Configuration#B, Scenario#A/Configuration#B.
* OPPO [8]
  + Proposal 11: Study the techniques of pre-processing at model input and post-processing at model output to enable the generalization capability of AI/ML model.
* Lenovo [15]
  + For evaluating the generalizability of an AI/ML model for beam management, the full list of network conditions/scenarios/parameter values need to be discussed and decided. Further, based on the effort and time required for testing an AI/ML model under such different network conditions/scenarios/parameter values, consider a limited set of parameters to for testing whether an AI/ML model is generalizable.
  + Generalizability of a proposed AI/ML model for beam management is evaluated by computing all the KPIs, inclusive of all the gains achieved and all the costs incurred, by the model for each of the different network conditions/scenarios/parameter values.
  + Discuss how to decide on the generalization ability of an AI/ML model based on the KPIs, inclusive of all the gains achieved and the costs incurred, that are evaluated for each of the different network conditions/scenarios/parameter values. Further, consider the threshold-based methods discussed above for further study.
* CAICT [16]
  + Proposal 1: The framework agreement in 9.2.2.1 for verifying the generalization performance of an AI/ML model could also be considered as a starting point for BM.

#### FL1: (closed) Principle for generalization performance evaluation

**Proposal 3-1-1a: (Same agreements as in 9.2.2.1)**

**The following cases are considered for verifying the generalization performance of an AI/ML model over various scenarios/configurations as a starting point:**

* **Case 1: The AI/ML model is trained based on training dataset from one Scenario#A/Configuration#A, and then the AI/ML model performs inference/test on a dataset from the same Scenario#A/Configuration#A**
* **Case 2: The AI/ML model is trained based on training dataset from one Scenario#A/Configuration#A, and then the AI/ML model performs inference/test on a different dataset than Scenario#A/Configuration#A, e.g., Scenario#B/Configuration#B, Scenario#A/Configuration#B**
* **Case 3: The AI/ML model is trained based on training dataset constructed by mixing datasets from multiple scenarios/configurations including Scenario#A/Configuration#A and a different dataset than Scenario#A/Configuration#A, e.g., Scenario#B/Configuration#B, Scenario#A/Configuration#B, and then the AI/ML model performs inference/test on a dataset from a single Scenario/Configuration from the multiple scenarios/configurations, e.g., Scenario#A/Configuration#A, Scenario#B/Configuration#B, Scenario#A/Configuration#B.**
  + - * **Note: Companies to report the ratio for dataset mixing**
      * **Note: number of the multiple scenarios/configurations can be larger than two**
* **FFS the detailed set of scenarios/configurations**
* **FFS other cases for generalization verification, e.g.,**
  + - * **Case 2A: The AI/ML model is trained based on training dataset from one Scenario#A/Configuration#A, and then the AI/ML model is updated based on a fine-tuning dataset different than Scenario#A/Configuration#A, e.g., Scenario#B/Configuration#B, Scenario#A/Configuration#B. After that, the AI/ML model is tested on a different dataset than Scenario#A/Configuration#A, e.g., subject to Scenario#B/Configuration#B, Scenario#A/Configuration#B.**

|  |  |
| --- | --- |
| Supporting companies | Futurewei, Google, MediaTek, HwHiSi, OPPO, Spreadtrum,NTT DOCOMO, Lenovo (Pl. See comments), Ericsson, vivo, Samsung, CATT, ZTE, NVIDIA |
| Objecting companies |  |

**Please provide your view Proposal 3-1-1a:**

|  |  |
| --- | --- |
| Company | Comments |
| LG | We are ok in general, but do we really need to make this agreement? During the discussion in 9.2.2.1, as Rapporteur mentioned, this issue is more related to general framework. So, making simple conclusion/guideline seems sufficient. |
| Xiaomi | We prefer to consider case 2 with higher priority than case 2A. |
| Nokia | Ok. it is not required to list the last sub-bullet. |
| Lenovo | The proposal is OK in general. *However*, as we expressed our views in RAN1#109, for a given AI/ML model, we find its generalization ability by testing the performance of AI/ML model under different possible scenarios/configurations and/or under possible changes in the statistical properties of the data.  Further, we do not see a need to mention how the AI/ML model should be trained. There are different ways/techniques in the field of AI/ML to achieve generalization. While “training with mixed datasets” is one way to generalizability, we believe that companies should be allowed to explore and employ a training technique that they think is suitable. Companies can report how the proposed AI/ML model was trained (to help re-producibility of results).  No matter how an AI/ML model is trained, the main concern should be how the given AI/ML model performs across different scenarios/configurations so that we know its generalization capability. Hence, what are the different scenarios/configurations that need to be considered for testing the performance of for generalization (as being discussed under Proposal 3-1-2a) should be of more concern than how do we train the AI/ML model. |
| FL | Agreed as WA in GTW as **Proposal 3-1-1a** |

**Proposed scenarios/configurations for generalization**

* Huawei/HiSi [2]:
  + Proposal 9: For verifying the AI/ML model generalization for spatial domain beam management, the scenarios/configurations for performing the inference for the AI/ML model should initially consider the following aspects:
    - Various channel types, e.g., UMa, UMi, InH
    - Various numbers of beams in Set A (including Tx beams and/or Rx beams)
    - Various Tx beam widths of Set B, e.g., wide beam, narrow beam
    - Various numbers of Set B (including Tx beams and/or Rx beams)
    - Various patterns of Set B, if Set B is a subset of Set A
  + Proposal 12: For verifying the AI/ML model generalization for temporal domain beam prediction, the scenarios/configurations for performing the inference for the AI/ML model should initially consider the following aspects:
    - Various channel types, e.g., UMa, UMi, InH
    - Various numbers of beams in Set A (including Tx beams and/or Rx beams)
    - Various Tx beam widths of Set B, e.g., wide beam, narrow beam
    - Various numbers of Set B (including Tx beams and/or Rx beams)
    - Various patterns of Set B, if Set B is a subset of Set A
    - Various UE speeds (e.g., 30km/h, 60km/h, 90km/h, 120km/h)
    - Various types of UE trajectories (e.g., Option 2/3/4)
* ZTE [3]
  + Proposal 4: Different inputs of AI/ML model (number/pattern of beams (pairs) in Set B, etc) can be considered for the evaluation of model generalization capability.
  + Proposal 6: Different UE speeds can be considered for the evaluation of model generalization capability for temporal beam prediction.
  + Proposal 6: Different UE speeds can be considered for the evaluation of model generalization capability for temporal beam prediction.
* Vivo [5]
  + Proposal 11: To study and evaluate generalization, at least the aspects including different scenarios, different UE speeds, different number of Tx beams and Rx beams, and different gNB/UE antenna configurations, should be prioritized.
  + Proposal 12: For evaluation of generalization performance, support to evaluate KPIs for a separately generated testing dataset generation method with at least 1 target parameter difference. Multiple target parameters can also be verified in further study.
  + ***Beam pair prediction with expected beam information***
  + Proposal 17: Study generalization performance of different number of Tx/Rx beams in BM-Case1.
  + Proposal 18: Study beam pair prediction with expected information as the AI input as one of the solutions for generalization to different number of Tx/Rx beams in BM-Case1.
  + Proposal 19: Further study expected information method in BM-Case2.
  + Proposal 20: Further study multiple expected beam information simultaneously used in AI input.
  + ***Generalization study for different beam shape patterns***
  + Proposal 22: Further study generalization performance for different antenna configurations and different beam shapes in BM-Case1.
  + Proposal 23: Further study assistance information, such as beam shape pattern, 3dB beam width, etc., as model input to address performance deterioration for generalization of different beam shapes in BM-Case1.
  + ***BM-Case 2***
  + Proposal 24: Further study beam pair prediction scheme with expected information as AI input for improving generalization performance in BM-Case2.
  + ***BM-Case 2: with different beam shape patterns***
  + Proposal 26: Further study generalization performance for different antenna configurations and different beam shapes in BM-Case2.
  + Proposal 27: Further study assistance information, such as beam shape pattern, 3dB beam width, etc., as model input to address performance deterioration for generalization of different beam shapes in BM-Case2.
  + Proposal 28: Suggest to use beam pointing angle or global beam ID as assistance information for AI model input.
* China Telecom [7]
  + Proposal 4: For beam management enhancement evaluations, to verify the generalization performance of an AI/ML model over various scenarios, the set of scenarios are considered focusing on one or more of the following aspects as a starting point:
    - Various deployment scenarios (e.g., UMa, UMi, InH)
    - Various outdoor/indoor UE distributions for UMa/UMi
    - Various carrier frequencies
    - Other aspects of scenarios are not precluded, e.g., various antenna spacing, various antenna virtualization (TxRU mapping), various ISDs, various UE speeds, etc.
* Ericsson [11]
  + Proposal 5: Discuss generalization in terms of different UE parameters, NW settings, deployment scenarios and propagation environment scenarios as a starting point
* CATT [12]
  + Proposal 3: For AI/ML based beam management, the AI/ML model trained with mixed numbers of beams (pairs) in Set B has significant generalization performance for different numbers of beams (pairs) in Set B.
* CAICT [16]
  + Proposal 2: For verifying the generalization performance of an AI/ML model, different NW settings and UE parameters could be classified as configurations.
* Xiaomi [17]
  + Proposal 5: To investigate the model generalization capability, the following aspect(s) can be considered with high priority for the evaluation for AI/ML in beam management:
    - Different UE parameters: UE speed, number of Rx beam
    - Different Scenarios, UMa, UMi including UE distribution, etc
* Nokia [19]
  + Proposal 9: RAN1 further investigates the ML model generalization capabilities of BM Case-1 with configurations that can be either included or not included in the training dataset. The following list of configurations can be prioritized:
    - gNB antenna array dimensions, e.g., 4x8 and 8x16.
    - UE antenna array dimensions/Number of Panels.
    - Set A dimension, e.g., 64 beams and 128 beams.
    - Set B dimension, e.g., 16 beams and 32 beams.
  + Proposal 10: RAN1 further investigates the ML model generalization capabilities of BM Case-1 with scenarios that can be either included or not included in the training dataset. The following list of scenarios can be prioritized:
    - Channel propagation models, e.g. UMa/UMi.
    - Outdoor/Indoor UE distribution (e.g. 100% Oudoor, 80% Outdoor/20% Indoor).
* Apple [21]
  + Proposal 2: For AI model generalization, discuss aspects related to analog beam design, antenna configurations including M/N, and antenna spacing and deployment scenario.
  + Observation: The AI/ML model trained with Dataset 1 does not generalize well to Dataset 2:
    - where
      * Dataset 1 is with d\_V=0.5,d\_H=0.5.
      * Dataset 2 is with d\_V=0.8,d\_H=0.4.
    - With mismatched AI model, the AI performance with set B beam at 16 beams is worse than the AI performance with set B at 8 beams without mismatched AI model.
* NVIDIA [23]
  + Proposal 2: To investigate the model generalization capability, the following aspects can be considered for the evaluation for AI/ML in beam management:
    - Different UE parameters: UE speed, UE antenna configuration, UE trajectory, number of Rx beams, UE antenna height, etc.
    - Different NW settings: BS antenna configuration (e.g., number of Tx beams, Tx beam width, TX beam boresight directions, etc.), Tx beam pattern, BS antenna height, etc.
    - Different Scenarios: UMa, UMi, including UE distribution, etc.
* Samsung [24]
  + Proposal # 8: Generalization is defined for UE side AI/ML model and gNB side AI/ML model separately.
  + Proposal # 9: For UE side AI/ML model, the following can be considered to verify the generalization performance.
    - Different UE parameters: UE speed, UE trajectories
    - Different gNB setting: number of Tx beam, Tx beam widths, Tx beam pattern, number or pattern in Set B (when applicable),
    - Different Scenario, like UMa, UMi including UE distribution
  + Proposal # 10: For gNB side AI/ML model, the following can be considered as a starting point to verify the generalization performance.
    - Different UE parameters: e.g., UE trajectories, UE speed, UE antenna config, number of Rx beam (when applicable),
    - FFS Different gNB setting: number of Tx beam, different beam widths, Tx beam pattern, number or pattern in Set B(when applicable)
    - FFS Scenario, like UMa, UMi including UE distribution e.g., outdoor: in door
* Qualcomm [26]
  + Proposal 3: Consider the following categorizations for definition of scenarios/configurations for evaluating the generalization capability of AI/ML models for temporal beam prediction:
    - Inter-site (heterogeneous): train AI/ML model on a first set of deployment type(s) and test it on a second (unseen) deployment type.
    - Inter-site (homogeneous): train on a first set of site(s) of a given deployment type and test it on a second (unseen) site of that same deployment type.
    - Intra-site: train AI/ML model for a given site and test it on unseen variations within that same site.
    - Across configurations: train AI/ML model on a first set of configuration(s) and test on a second configuration

#### FL3: Assumptions for generalization performance verification

**Proposal 3-1-2a:**

* **For BM Case-1 and BM Case 2, to verify the generalization performance of an AI/ML model over various scenarios, the set of scenarios are considered focusing on one or more of the following aspects as a starting point:**
  + **Various deployment scenarios (e.g., UMa, UMi, InH; e.g., ISD 200m, ISD 500m)**
  + **Various outdoor/indoor UE distributions (e.g., 0:1, 2:8, 5:5, 8:2, 1:0)**
  + **Various UE speeds (e.g., 3km/h, 30km/h, 60km/h, 90km/h, 120km/h)**
  + **Various UE parameters (e.g., number of Rx beam: 4, 8)**
  + **Various gNB settings (e.g., number of Tx beam: 32, 64)**
  + **Other aspects of scenarios are not precluded, e.g., various number of Set B of beam(pairs) (e.g., ¼, 1/8 of set A beams (pairs)), carrier frequencies, etc.**
  + **Companies to report the selected scenarios for generalization verification**

|  |  |
| --- | --- |
| Supporting companies | Google, Xiaomi,NTT DOCOMO, vivo, Samsung, NVIDIA, Futurewei. CEWiT, CAICT, LG, MediaTek |
| Objecting companies |  |

**Please provide your view Proposal 3-1-2a, and your views on the following questions:**

Whether the set of scenarios for generalization performance verification needs to consider:

* A: BM Case-1 and BM Case-2
* B: AI model inference node, e.g. @UE side vs @ gNB side
* C: Different cases for evaluation: e.g., DL Tx beam prediction, DL Rx beam prediction, Tx-Rx beam pair prediction
* D: Others

|  |  |
| --- | --- |
| Company | Comments |
| Google | No to A/B/C |
| vivo | Different sets of generalization parameters need to consider different cases in A/B/C. For example, UE speeds need to be considered for BM Case 2, but not for BM Case 1. Rx beam numbers need to considered for Tx/Rx beam pair or Rx beam prediction, but probably not for Tx beam prediction. Although it makes sense to consider these, it may not be needed to list all the details in an example. We can generally add a bullet and let the companies report further details in their evaluations. |
| FL1 | Please continue the discussion. |
| HW/HiSi | Our views on proposal 3-1-2a:  We prefer to not give examples in brackets for now, but let companies report. There is no need for these examples since we already have the agreed simulation assumptions. And if companies want to additionally test other values, this can be reported.  The second last sub-bullet, in our understanding, *(“****Other aspects of scenarios are not precluded, e.g., various number of Set B of beam(pairs) (e.g., ¼, 1/8 of set A beams (pairs)), carrier frequencies, etc.***”) does not really seem to fit into this proposal  On A/B/C  A: Generalization could be different for BM-Case 1 and BM-Case 2, for example UE speeds should be included in Case 2, but for Case 1 this does not really seem needed-  B:Yes, we should at least discuss whether UE side and gNB side models have different generalization studies.  C: We think this can be postponed until we have progressed on the DL Tx, DL Rx, Tx-Rx beam pair prediction. |
| Futurewei | We believe most of the scenarios/settings are applicable for either BM-Case1 or BM-Case2. However, “Various UE speeds” is more relevant to BM-Case2. |
| Qualcomm | For UE and gNB parameters, codebook size has been mentioned, but also suggest adding other codebook characteristics such as beam boresight directions, 3dB beamwidth, etc. |
| Apple | “base station antenna configuration” needs to be added to the list. As shown in our study, the antenna element spacing d\_H/d\_v is important as it impact generalization performance. To make it more general, we can mention “d\_H/d\_v, d\_{H,g}/d\_{V,g}” |
| Spreadtrum | Our views on A/B/C  We can consider both A and B for the generalization. C should be discussed with lower priority. |
| CAICT | Share the same view as HW. |
| NTT DOCOMO | Good to consider A/B/C in generalization evaluation.  Companies could report their selection for A/B/C respectively, when they report the generalization evaluation results. |
| OPPO | We think one of the sub-bullet could be updated as   * + **Various UE parameters (e.g., number of Rx beam per panel: 4, 8)**   In addition, for FR2, if ISD 500m is adopted, we worry the performance on L1-RSRP and Tput would be very low given our experience of SLS simulation.  A: Yes. BM-Case2 focuses on the scenario with mobility (speed > = 30km/h) and all outdoor UEs.  B: No. Assuming offline trained model, no matter it is deployed at either NW or UE, the generalization performance can be evaluated without difference. But anything we missed, please feel free to enlighten. Thanks.  C: Since DL Rx beam prediction is considered with less spec impact, we suggest to focus on the other two cases of beam prediction, i.e. DL Tx-Rx beam pair and DL Tx beam. |
| Nokia | C is not required. A and B are fine. |
| LG | A/B: Ok to consider  C: it can be postponed. |
| CMCC | Different sub use case can consider different aspects for generalization evaluation. |
| ZTE | Both A and B can be considered and C can be postponed. Besides, we prefer BM Case-1 to be used in scenarios with low UE speeds and thus UE speeds don't need to be considered for generalization performance verification of BM-Case1. Otherwise, its performance may be degraded and it will be inferior to the temporal beam prediction or spatial-time domain beam prediction. |
| FL2 | Please continue to comment in the following updated proposal.  **Proposal 3-1-2b:**   * **For BM Case-1 and BM Case 2, to verify the generalization performance of an AI/ML model over various scenarios/configurations, the set of scenarios/configurations are considered focusing on one or more of the following aspects as a starting point:**   + **Various deployment scenarios**   + **Various outdoor/indoor UE distributions**   + **Various UE speeds**   + **Various UE parameters**   + **Various gNB settings**   + **Other aspects of scenarios/configurations are not precluded**   + **The selected scenarios/configurations for generalization verification may consider the AI model inference node (e.g., @UE or @gNB) and use case (e.g., BM-Case1, or BM-Case2)**   + **Companies to report the selected scenarios/configurations for generalization verification** |
| Lenovo | We support the proposal.  A: We believe we need to agree on different scenarios to be considered for evaluating generalization for both BM case-1 and BM case-2.  B: Different scenarios/configurations to be considered for generalization may not vary depending on where the AI/ML model is deployed. Thus, we don’t think this is important to be considered.  C: Both,   1. Joint Tx-Rx beam pair prediction 2. Prediction of Tx beam and Rx beams separately   may result in very different performance. If the AI/ML model is capable of only one type of prediction (i.e., either (a) or (b)) then it needs to evaluate generalization for only one case. Otherwise, it needs to report generalization ability for both (a) and (b). |
| CATT | For A, we support generalization performance verification needs to consider BM Case-1 and BM Case-2.  For B, we don’t think the generalization performance is related with AI model inference node based on our simulation results.  For C, we think the generalization performance can be related with different cases for evaluation. But there is a discussion in agenda 9.2.3.2 whether to down select DL Tx beam prediction, DL Rx beam prediction, Tx-Rx beam pair prediction. Most companies prefer to de-prioritize DL Rx beam prediction. Maybe we can wait and add FFS in the proposal.  Thus, we suggest to update the proposal Proposal 3-1-2b as following:  **Update: Proposal 3-1-2b:**   * **For BM Case-1 and BM Case 2, to verify the generalization performance of an AI/ML model over various scenarios/configurations, the set of scenarios/configurations are considered focusing on one or more of the following aspects as a starting point:**   + **Various deployment scenarios**   + **Various outdoor/indoor UE distributions**   + **Various UE speeds**   + **Various UE parameters**   + **Various gNB settings**   + **Other aspects of scenarios/configurations are not precluded**   + **The selected scenarios/configurations for generalization verification may consider the ~~AI model inference node (e.g., @UE or @gNB) and~~ use case (e.g., BM-Case1, or BM-Case2)**     - **FFS: Different cases for evaluation: e.g., DL Tx beam prediction, DL Rx beam prediction, Tx-Rx beam pair prediction**   + **Companies to report the selected scenarios/configurations for generalization verification**   FL3: We say “may consider”. Whether to consider it up to companies. |
| MediaTek | We support the proposal. |
| Futurewei | We are ok with Proposal 3-1-2b in general. Regarding question C, we believe generalization is needed for all cases while companies may decide generalization scenarios/configurations that are applicable for the beam prediction case discussed in their study. |
| Intel | Ok with the proposal in general. Suggest to move the red sub-bullet to the same level as the main bullet. |
| Intel | Ok with the proposal in general. Suggest to move the red sub-bullet to the same level as the main bullet.  FL3: Updated |
| CAICT | We support the proposal and think it can apply to A/B/C. |
| Apple | “various gNB settings” in our view should include CSI-RS resources/resource set configurations, analog beam design, including antenna spacing, … If it is common understanding all the listed items are covered by “various gNB settings”, we don’t need to itemize them. Otherwise, some clarification is needed. |
| vivo | OK with the updated proposal from FL |
| ZTE | We think it may be helpful to provide some typical candidates in each sub-bullet as in Proposal 3-1-2a for better alignment. |
| Samsung | We support FL’s proposal 3-1-2b**.** |
| Xiaomi | Prefer to make these two sub-bullet more detail   * + **Various UE parameters**      - **E.g., Number of Rx beam**   + **Various gNB settings**      - **E.g., Number of Tx beam**   FL3: There are other proposals from other companies, to be fair, I will not list examples. Companies can report their assumptions. |
| Ericsson | We think it would be useful to clarify what is a configuration and a scenario. Our proposal is to clarify this by using the following update marked in blue. It is further proposed to generalize the UE speed to also cover different UE trajectory, rotation, speed etc.   * **For BM Case-1 and BM Case 2, to verify the generalization performance of an AI/ML model over various scenarios/configurations, the set of scenarios/configurations are considered focusing on one or more of the following aspects as a starting point:**   + **Scenarios**     - **Various deployment scenarios**     - **Various outdoor/indoor UE distributions**     - **Various UE mobility ~~speeds~~**   + **Configurations**     - **Various UE parameters**     - **Various gNB settings**   + **Other aspects of scenarios/configurations are not precluded**   + **The selected scenarios/configurations for generalization verification may consider the AI model inference node (e.g., @UE or @gNB) and use case (e.g., BM-Case1, or BM-Case2)**   **Companies to report the selected scenarios/configurations for generalization verification**  FL3: we use UE speed in the simulation assumptions. |
| qualcomm | For “various deployment scenarios” we believe some elaboration is needed. One example is train for UMi deployment, test in (unseen) UMa deployment (heterogeneous inter-site). Another example is to train in a first set of site(s) of a given deployment type (e.g., UMi) and test it on a second (unseen) site of that *same* deployment type (homogeneous inter-site). Also, we believe a few examples (at least in the form of FFS) could help elaborate what is meant by “Various UE parameters” and “Various gNB settings”, e.g., UE/gNB codebook, etc.  FL3: for the examples of various deployment scenarios, I think both can be included depends on company reports. |
| Fujitsu | We support the proposal 3-1-2b in general. |
| LG | Support the proposal. |
| NTT DOCOMO | We support the proposal. |
| HW/HiSi | We support the updated proposal 3-1-2b. |
| CMCC | We support the proposal. |
| FL3 | In principle, I tried not to list examples. However, if we change speed to mobility, I feel it may not be clear on what it refers. Companies are encouraged to check whether the examples for UE mobility is needed or not.  **Proposal 3-1-2c:**   * **For BM Case-1 and BM Case 2, to verify the generalization performance of an AI/ML model over various scenarios/configurations, the set of scenarios/configurations are considered focusing on one or more of the following aspects as a starting point:**   + **Scenarios**     - **Various deployment scenarios**     - **Various outdoor/indoor UE distributions**     - **Various UE mobility [e.g., UE trajectory, rotation, speed etc.]**   + **Configurations**     - **Various UE parameters**     - **Various gNB settings**   + **Other aspects of scenarios/configurations are not precluded**   + **The selected scenarios/configurations for generalization verification may consider the AI model inference node (e.g., @UE or @gNB) and use case (e.g., BM-Case1, or BM-Case2)** * **Companies to report the selected scenarios/configurations for generalization verification** |
| Xiaomi | We are OK with the proposal |
| ZTE | Since variable set B has been agreed to be studied in RAN1#110 and is being discussed in Proposal 4-3-1c as an important aspect of beam pattern selection, we suggest to add various Set B of beam(pairs) for generalization evaluation.   * + **Configurations**     - **Various UE parameters**     - **Various gNB settings**     - **Various Set B of beam(pairs)** |
| OPPO | Fine with the proposal.  Look at the other two sub-bullets under scenarios, there are no examples. It seems clear what UE mobility refers to. Moreover, UE speed > 3km/h may only applicable for BM-Case2, so it looks neat to remove the e.g. part. |
| HW/HiSi | Support |
| CATT | Fine with the proposal. Prefer to remove the e.g. part for mobility, if there is a common understanding. |
| Samsung | We are fine with the Proposal 3-1-2c. |
| LG | Fine with the proposal. Also, we agree with OPPO for removing “e.g.” for UE mobility. We don’t down-select the UE trajectory model yet, but generalization for various UE trajectory models (if I understood correctly) seems not needed. |
| Lenovo | Support proposal 3-2-1c. |
| Ericsson | Support |
| MediaTek | Support the latest proposal 3-1-2c, prefer to keep the examples for UE mobility. |

## 3.2 (on hold) Evaluation Results

**Observations for generalization**:

* Futurewei [1]
  + Observation 5: In AI/ML model generalization across different scenarios for spatial-domain beam prediction when using fixed beam pattern sampling, our experiments show the following based on the datasets we used:
    - The AI/ML model trained using dataset generated for UMi\_UMa scenario/channel model CAN generalize to UMi\_UMi scenario/channel model without performance degradation.
    - The AI/ML model trained using dataset generated for UMi\_UMi scenario/channel model may be used for UMi\_UMa scenario/channel model with small performance degradation.
  + Observation 6: In AI/ML model generalization across different scenarios for spatial-domain beam prediction when using pre-configured beam patterns sampling, our experiments show the following based on the datasets we used:
    - The AI/ML model trained using dataset generated for UMi\_UMa scenario/channel model CAN generalize to UMi\_UMi scenario/channel model without performance degradation.
    - The AI/ML model trained using dataset generated for UMi\_UMi scenario/channel model may be used for UMi\_UMa scenario/channel model with small performance degradation
* ZTE [3]
  + Observation 4:The case of AI generalization with different model inputs can achieve a better performance than that of the case of <T8 R1>, but is outperformed by the case of <T32 R1>.
  + Observation 7: The AI/ML method suffers from a little performance loss for scenarios with mixed UE speeds.
* Vivo [5]
  + ***DL Tx beam prediction***
  + Observation 10: More flexible AI model deployment for different number of Rx beams can be obtained by expected Rx beam information method with only marginal performance loss as well as expected Tx beam information scheme.
  + ***Generalization study for different beam shape patterns***
  + Observation 14: As the difference of beam shape pattern increases, the performance loss of both average RSRP difference and beam prediction accuracy increases along with the difference of the antenna configurations between training subset and validation subset.
  + ***BM-Case 2: with different beam shape patterns***
  + Observation 17: Performance loss can be observed with difference datasets represented different beam shape patterns for training and validation in BM-Case2.
  + Observation 18: For the case using local beam ID as model input, beam loss and accuracy degenerate significantly compared to the performance of AI model training and inference with beam pointing angle.
* OPPO [8]
  + Observation 12: Thanks to generalization capability of well-trained AI/ML model, changing scenario from UMa to UMi may not necessarily deteriorate the beam prediction performance.
  + Observation 13: Changing beam pair configuration on Set B and Set A from training phase to inference phase would slightly lower the beam prediction performance.
  + Observation 14: When more predicted beam pairs are provided by AI/ML model, e.g. Top-K = 4, the beam selection accuracy can be up to 95% and avg. L1-RSRP difference can be lower than 1dB.
* Ericsson [11]
  + Observation 9: With identical antenna configuration, initial evaluations indicates that a model trained in one cell is found to be generalized well to another cell except for one or two fixed cells.
* CATT [12]
  + Observation 5: For AI/ML based beam management, there is significant performance degradation when the number of beams (pairs) of Set B for inference is smaller than that of Set B for training.
  + Observation 6: For AI/ML based beam management, the performance of AI/ML model, which is trained with mixed numbers of beams (pairs) in Set B, is similarity for different numbers of beams (pairs) in Set B.
* Fujitsu [13]
  + Observation 1: The module of pre-processing provides the flexibility to adapt to variable Set B for AI/ML model.
  + Observation 3: For variable Set B, the model trained by mixed dataset constructed by samples with different sampling rates from beams of Set A will improve the performance of AI/ML model.
  + Observation 4: For variable Set B, the performance of the model trained by mixed dataset is almost the same as the model trained by separated dataset for fixed Set B.
* Xiaomi [17]
  + BM Case 1:
  + Observation 2: AI model trained by hybrid data of Uma and Umi for beam prediction in spatial domain can provide good generalization capability for Uma or Umi. While AI model trained by data of only Uma or only Umi provide a little worse generalization capability
  + Observation 3: AI model trained by hybrid data of different UE distribution for beam prediction in spatial domain can provide good generalization capability. While AI model trained by data of only UE distribution Option A provides a little worse generalization capability for UE distribution Option B.
  + Observation 4: AI model for beam prediction in spatial domain can provide good generalization capability among different number of UE Rx beam, e.g., AI model with more Rx beam number can be applied for beam prediction of less Rx beam number.
  + BM Case 2:
  + Observation 8: AI model for beam prediction in time domain trained by data of 30km/h or only 60 km/h or hybrid can provide good generalization capability to UE speed with both 60km/h and 30km/h.
* Nokia [19]
  + Observation 11: model's generalization capabilities should be assessed considering different combination of configurations as ML model performances can be affected significantly.
  + Observation 12: Several configurations/scenarios can be considered for assessing the ML model generalization capabilities. In this study, we considered the gNB antenna array dimensions, but other configurations/scenarios are not precluded. Supporting multiple configurations may further affect ML performance.
  + Observation 13: For BM-Case1, the Set A/B model generalization issue can be addressed with a training model based on an oversampled Set C that satisfies Set B∈Set A∈Set C for any given Set A/B.
  + Observation 14: For BM Case1, the training model with a fixed Set B pattern will have poor beam prediction performance if the testing Set B does not match with the training Set B.
  + Observation 15: For BM-Case1, training the model with random Set B is possible to provide beam prediction performance close to the optimal case – training and testing on the same Set B.
* Apple [21]
  + Observation: The AI/ML model trained with Dataset 1 does not generalize well to Dataset 2:
    - where
      * Dataset 1 is with d\_V=0.5,d\_H=0.5.
      * Dataset 2 is with d\_V=0.8,d\_H=0.4.
    - With mismatched AI model, the AI performance with set B beam at 16 beams is worse than the AI performance with set B at 8 beams without mismatched AI model.
* Samsung [24]
  + Observation # 12: For DL TX beam prediction and beam pair prediction, AI/ML model performs the best when the training and testing dataset are drawn from the same UE speed. However, performance degradation is observed when the training dataset and testing datasets are drawn from different UE speed.
  + Observation # 13: For DL TX beam prediction and beam pair prediction, training a model with a mixture of dataset drawn from a range of UE speeds allows the model to perform well over a range of UE speeds.

# AI/ML related assumptions

## 4.1 (on hold) Inputs of AI/ML models

In RAN 1 #110, the following alternatives were agreed.

|  |
| --- |
| **Agreement**  *For* the sub use case BM-Case1 and BM-Case2, further study the following alternatives for the predicted beams:   * Alt.1: DL Tx beam prediction * Alt.2: DL Rx beam prediction * Alt.3: Beam pair prediction (a beam pair consists of a DL Tx beam and a corresponding DL Rx beam) * Note1: DL Rx beam prediction may or may not have spec impact |

In this section, further discussion on inputs for each alternative. The following observations/proposals were provided in contributions:

* Huawei [2]
  + Observation 1: For the AI/ML-based beam prediction mechanism, Option 2 (DL Tx beam prediction) may also achieve best Tx-Rx beam combination by DL Tx beam prediction and legacy Rx beam sweeping.
  + Proposal 1: For the evaluation of AI/ML-based beam prediction mechanism,
    - Option 2 (DL Tx beam prediction) should be considered as the starting point.
      * Both Case A (best Rx beam) and Case B (same specific Rx beam) can be adopted and reported by companies
    - Option 1 (Tx-Rx beam pair prediction) can be also evaluated to justify the additional performance gain over Option 2.
    - Option 3 (DL Rx beam prediction) can be considered with lower priority.
* ZTE [3]
  + Observation 3: With a same sampling rate on the whole beam space, the Tx beam prediction obtains a better performance than that of the Tx-Rx beam pairs prediction.
  + Proposal 5: Whether to choose Alt 1 or Alt 4 needs further discussion according to the beam pattern selection.
    - *Note by FL0: Alt 1: only RSRP; Alt 4: RSRP and beam IDs*
* Vivo [5]
  + *DL Tx beam prediction*
  + Observation 11: Significant performance deterioration can be observed in two-step beam prediction with non-best Rx beam, even for the 2nd best Rx beam.
  + Observation 12: The performance of two-step beam prediction with the best Rx beam provides considerable improvement, as decreased prediction difficulty from predicting 256 beam pairs to 32 beam pairs by acquiring precise best Rx beam of each sample.
  + Observation 13: Large performance deterioration can be observed if the Rx beam assumptions of training and inference are different for DL Tx beam prediction scheme.
  + Proposal 21: Study DL Tx beam prediction with different Rx beam assumptions as one of the solutions for generalization to different number of Tx/Rx beams in BM-Case1.
* OPPO [8]
  + For BM-Case1 and BM-Case2, at least support beam pair prediction (Alt.3) as the key feature of representative sub use cases.
  + Observation 1 The input of AI/ML model for beam prediction are element-wise sensitive, therefore the L1-RSRPs of Tx and/or Rx beams in Set B should be input in proper order.
  + Proposal 3: For BM-Case1 and BM-Case2, suggest to adopt L1-RSRP measurement based on Set B as input of AI/ML model.
  + FL0: can be discussed in 9.2.3.2
* CATT [12]
  + *Beam pair prediction:*
  + Observation 2: For Beam pair prediction, additional Beam ID input have significant performance gain compared with beam prediction accuracy with random pattern using L1-RSRP input only.
  + *DL Tx beam prediction:*
  + Observation 3: For DL Tx beam prediction, beam prediction accuracy with fixed pattern has better performance than random pattern, since beam ID is implicit in the fixed pattern.
  + Observation 4: For DL Tx beam prediction, additional Beam ID input have significant performance gain compared with beam prediction accuracy with random pattern using L1-RSRP input only.
  + Proposal 2: For AI/ML based beam management, DL Tx and/or Rx beam ID is supported as an additional input besides L1-RSRP measurement based on Set B.
  + FL0: Some clarification on inputs can be discussed in 9.2.3.2
* Xiaomi [17]
  + Observation 1: AI based beam prediction in spatial domain can provide good performance. And the performance can be further improved by inputting corresponding beam pair ID in addition to measured L1-RSRP or by inputting L1-RSRP of same beam pair IDs.
  + FL0: Some clarification on inputs can be discussed in 9.2.3.2
* Mediatek [20]:
  + Proposal 9: For AI/ML-based spatial domain beam prediction evaluation, adopt the RSRP of beams in Set B as the AI/ML model inputs. Additional information to the input of AI/ML model is not excluded.
  + Proposal 10: Adopt one of the following as the output of AI/ML model: (i) beam index of highest RSRP Set A of beams. (ii) RSRPs of all the Set A of beams.
* Samsung [24]
  + Proposal # 3: Deprioritize the study of Rx beam prediction in this study item for AI/ML in beam management.
  + Proposal # 4: At least for BM Case 1, the following options can be further studied and potential down selection as the inputs of AI model:
    - Option 1: For Tx-Rx beam pair prediction:
      * L1-RSRP measurements of Tx-Rx beam pairs in Set B
        + FFS on the selection of Tx-Rx beam pairs in Set B
    - Option 2: For DL Tx beam prediction
      * L1-RSRP measurements of Tx beams in Set B, measured by one or multiple Rx beam(s), FFS:
        + The Rx beam is “best” Rx beam based on historical measurements
        + The Rx beam(s) is by UE implementation FFS fixed Rx beam or different Rx beam for measuring different Tx beams in Set B
        + FFS: The Rx beam(s) is fixed and configured by gNB or chosen by UE implementation
      * FFS on the number of Rx beams
    - FFS on other information as AI inputs
  + Proposal # 5: For AI inference at gNB side, DL Tx beam prediction is prioritized, and focus on the L1-RSRP measurements of Tx beams in Set B with explicit or implicit Tx beam ID as AI inputs.
* Qualcomm [26]:
  + Proposal 7: For both spatial and temporal prediction evaluation, consider the following options as inputs to AI/ML models for the study and potential down selection:
    - Option 1: For Tx-Rx beam pair prediction:
      * L1-RSRP of Tx-Rx beam pairs in Set B
    - Option 2: For DL Tx beam prediction
      * L1-RSRP of Tx beams in Set B, measured by a (set of) Rx beam(s) selected by UE
        + FFS on selection criteria of (set of) Rx beam(s) by UE
    - Option 3: For DL Rx beam prediction,
      * L1-RSRP of Rx beams in Set B (where Set B of beams is for Rx beam)
    - Note: DL Rx beam prediction may or may not have spec impact
    - Other inputs (e.g., CIR) are not preluded.
    - Note 1: Other assistance information is not precluded
    - Note 2: Options 1 and 3 are applicable to UE-side AI/ML models.

FL 0: Wait for the clarification in 9.2.3.2 on whether implicit beam ID is included or not, as well as some other discussions.

## 4.2 Number of beams Tx and/or Rx beams for Set A and Set B

* Huawei/HiSi [2]:
  + Proposal 5: For the evaluation of AI/ML-based beam management, for the construction of Set A, a DFT codebook with 64 DFT Tx beams and a denser codebook with 256 Tx beams should be considered.
  + Proposal 16: For the evaluation of beam prediction, RAN1 should study Set A with size of 64 and 256 beams to improve beam management related system performance, complexity and coverage over the legacy baseline.
  + Another issue is the number of beams in Set B. Since one of the main motivations to employ sparse beam sweeping is to save overhead compared to an exhaustive sweep, it can be considered to limit the number of beams in Set B relatively to the number of beams in Set A, e.g., number of beams in Set B should not exceed one fourth of the number of beams in Set A.
* Spreadtrum [4]:
  + Proposal 1: For both BM-Case1 and BM-Case2, unifying the number of beams contained in set A and set B should be considered.
* Vivo [5]
  + Proposal 1: Slightly prefer Alt 1 in evaluation assumption of Tx beam number, but we can live with Alt 2.
  + Proposal 2: Support 4 Rx beams per UE panel used at UE side for the evaluation of both temporal and spatial domain beam prediction.
* OPPO [8]
  + Observation 11: For spatial and temporal domain beam prediction, the longer gap between measurement on Set B and prediction among Set A, the deeper performance loss can be observed.
* LGE [10]
  + Proposal 3. It is preferred to fix the number of beams in Set A.
* Ericsson [11]
  + Observation 2: For NW-sided model, the variable number of beams could be due to UE only reporting a subset of the measured beams.
  + Proposal 4: Define the number of beams in set B as a fraction of beams in set A
* Nokia [19]
  + Proposal 1: For BM-Case1, given the current agreed NW antenna configuration, the number of DL Tx beams in Set A should be 32 or 64.
  + Observation 1: For BM-Case1, a large number of beams in Set B (e.g., 32) may not improve the prediction accuracy and the system throughput. Therefore, ML-based beam selection should consider a Set B with a maximum of 16 beams when Set A has 64 beams, hence Set B should have a max of ¼ of Set A beams.
  + Observation 2: For BM-Case1, a “sparse” Set B, or a random Set B pattern design, may cause throughput loss, especially for the cell-edge UE.
  + Observation 3: For BM-Case1, Set B RSRP may not be sufficient for beam prediction input in certain cases.
  + Proposal 2: For BM-Case1, RAN1 further study the case of Set A/B are DL Tx and Set B is a subset of Set A.
    - When Set B is a subset of Set A, RAN1 should consider a Set B with a maximum number of DL Tx beams that is ¼ of Set A beams.
  + Observation 4: For BM-Case1, the ML model using as input only RSRP measurements has performances that reduce significantly changing the number of RSRP measurements from 8 to 4, i.e. further down sampling Set A, from a ratio of ¼ to a ratio of 1/8.
* CEWiT [27]
  + Observation 2: When the size of Set B is increased, the performance of the AI/ML model improves.

#### FL1: (Closed)Number of Tx and Rx beams

**Proposal 4-2-1a:**

* **Adopt the following proposals as working assumption:**
* **For the evaluation of both BM-Case1 and BM-Case 2, 32 or 64 [or 256] Tx beams are used at NW side.** 
  + **Other values are not precluded and can be reported by companies.**
* **For the evaluation of both BM-Case1 and BM-Case 2, 4 Rx beams per UE panel are used at UE side.** 
  + **Other values are not precluded and can be reported by companies.**

|  |  |
| --- | --- |
| Supporting companies | Futurewei, Google, MediaTek, LG, Xiaomi, OPPO, Spreadtrum,NTT DOCOMO, Lenovo, vivo, Sausung, CATT, ZTE, NVIDIA |
| Objecting companies |  |

**Please provide your view Proposal 4-2-1a, if any.**

|  |  |
| --- | --- |
| Company | Comments |
| FL0 | The above number of beams are used/proposed by majority of companies. I understand that some company may have different number in the assumption. The door is not closed.  In FL’s view, it is good to align the assumption for further evaluation results collection.  Generalization is discussed separately; other values are not precluded. |
| HW/HiSi | We think that the brackets around the 256 should be removed. The reason is that 256 beams can achieve performance improvements compared to 64 beams, without more sweeping overhead.  It would be a good outcome to show the future potential of AI when used for beam management in cells with large number of beams.  Proposed update alt1:   * **Adopt the following proposals as working assumption:** * **For the evaluation of both BM-Case1 and BM-Case 2, 32 or 64 ~~[~~or 256~~]~~ Tx beams are used at NW side.**    + **Other values are not precluded and can be reported by companies.** * **For the evaluation of both BM-Case1 and BM-Case 2, 4 Rx beams per UE panel are used at UE side.**    + **Other values are not precluded and can be reported by companies.**   Alternatively, if the goal is to limit the options for TX beams in order to compare results across companies, we think that 64 could be taken as baseline and 32 and 256 can be optional values.  Proposed update alt2:   * **Adopt the following proposals as working assumption:** * **For the evaluation of both BM-Case1 and BM-Case 2, ~~32 or~~ 64 ~~[or 256~~] Tx beams are used at NW side.**    + **Other values (e.g. 32 or 256) are not precluded and can be reported by companies.** * **For the evaluation of both BM-Case1 and BM-Case 2, 4 Rx beams per UE panel are used at UE side.**    + **Other values are not precluded and can be reported by companies.** |
| Nokia | Ok with the FL or HW version. |
| Ericsson | Support FL or the update from HW |
| Samsung | We support current proposal is good enough |
| FL1 | Agreed in GTW on Monday  **Agreement**   * **For the evaluation of both BM-Case1 and BM-Case2, 32 or 64 downlink Tx beams (maximum number of available beams) at NW side.**    + **Other values, e.g., 256, etc, are not precluded and can be reported by companies.** * **For the evaluation of both BM-Case1 and BM-Case2, 4 or 8 downlink Rx beams (maximum number of available beams) per UE panel at UE side.**    + **Other values, e.g., 16, etc, are not precluded and can be reported by companies.** |

#### FL2: (close)Number beams(pairs) in Set B

**Proposal 4-2-2a:**

* **For evaluation for DL Tx beam prediction in BM-Case1, when Set B is a subset of Set A, the number of beams(pairs) in Set B is no more than ¼ of beams (pairs) in Set A.**

|  |  |
| --- | --- |
| Supporting companies | NTT DOCOMO |
| Objecting companies | Google (Too early for this restriction), Lenovo, MediaTek |

**Please provide your view Proposal 4-2-1, if any. And the views of following questions:**

**A:** whether to provide list of number of beams, e.g., 1/8 or 1/4 of Set A beams

**B:** the ratio of Set B beam pairs and Set A beam pairs

|  |  |
| --- | --- |
| Company | Comments |
| Google | We think this requires more study given different options (L1-RSRP, L1-RSRP + assistant info, CIR) are still open. |
| vivo | We can let the companies to report the used ratio. Further evaluation is needed to understand the maximum value. |
| FL1 | Please continue the discussion. |
| HW/HiSi | We support Proposal 4-2-2a.  Is there a typo in the question on the proposal number (“**Please provide your view Proposal 4-2-1**”)? |
| NVIDIA | It does not appear necessary at this stage to impose this restriction. |
| Futurewei | We think specifying such constraint in the study is not needed. |
| Qualcomm | Do not see the necessity at this point. |
| Apple | Premature to agree the proposal. |
| Samsung | Good to have some guidance, but no need to restrict. |
| OPPO | If companies can report the beam (pair) selection of Set B and Set A, then the ratio can be calculated accordingly. By far, it seems premature to fix the ratio between Set B and Set A. |
| Nokia | Support based on the simulation results. Good to align on these at least for future evaluations. |
| LG | No need such restriction at this stage. |
| CMCC | Not needed. |
| Fujitsu | If the evaluation results are tried to be aligned, it’s better to provide list of number of beams. |
| ZTE | Support to provide some ratio candidates as guidance, but there is no necessity to align the ratio at the current stage. |
| Lenovo | We do NOT support it. It is not at all desirable to have such a restriction in this study. It does not matter how many beams are in Set B as they are accounted for in at least one of the KPIs and the AI/ML meodels using different sizes for Set B can be compared fairly using gains (such as beam prediction accuracy, throughput etc.) and the costs incurred by the AI/Ml model (such as the overhead, computational complexity etc.). Please also refer to our response to proposal 4-3-1a. |
| MediaTek | We don’t support restricting Set B sizes at this point, given that we are already discussing to include the RS overhead as a KPI. The definition of RS overhead should be able to capture the benefit of using lower size of Set B. |
| Intel | Not needed. Companies can report assumptions. |
| Ericsson | This is not needed at this stage, due to the agreement on a number of TX & RX beams. |
| CATT | For evaluation, we don’t think it’s necessary to restrict the Set size. Companies can report the assumptions when providing the simulation results. |
| FL3 | This discussion is closed in this meeting.  Companies can report the number of beams in Set B based on the agreed Tx/Rx beams |

## 4.3 Selection of Set B of beams(pairs)

RAN 1 #110 agreed three options of the selection of Set B of beams(pairs).

|  |
| --- |
| **Agreement**   * **Study the following options on the selection of Set B of beams (pairs)**    + **Option 1: Set B is fixed across training and inference**     - **FFS on the beams of Set B**   + **Option 2: Set B is variable (e.g., different beams (pairs) patterns in each report/measurement during training and/or inference)**      - **FFS on fixed or variable number of beams (pairs)**     - **FFS on the details**   + **Other options are not precluded.**   + **FFS on the number of beams (pairs) in Set B**   + **Note: This does not preclude the alternative that Set B is different from Set A.** |

The following were discussed or assumed in the assumption for the evaluation of AI for BM:

* Futurewei [1]
  + Option 1: Fixed Beam Pattern
    - In this option, a defined fixed beam pattern with M select beams out of all the available beam pairs is applied for all the input samples. In our experiment, we use even-space sampling to pick M beam pairs (M  {4, 8, 12, 16, 20, 24, 28, 32}) from the total 256 beam pairs.
  + Option 2: Random Beam Patterns
    - In this option, we randomly select M beam (M  {4, 8, 12, 16, 20, 24, 28, 32}) from all available beams as input for each sample.
  + Option 3: Pre-configured Beam Patterns
    - In this option, we pre-defined a set of N (N = 5) different beam patterns, each with M selected beam pairs (M  {4, 8, 12, 16, 20, 24, 28, 32}), then one of them will be randomly chosen as input for each sample.
  + Observation 1: The performance of spatial-domain beam prediction using either fixed beam pattern sampling or pre-configured beam patterns sampling is improving when the training dataset size is increasing.
  + Observation 3: Performance of pre-configured beam patterns sampling is more sensitive to training dataset size increase compared to fixed beam pattern sampling.
  + Proposal 1: For AI/ML based spatial beam prediction, when using pre-configured beam patterns sampling approach, further study the trade-off between training dataset size and performance.
* Huawei/HiSi [2]:
  + Observation 2: For the selection of Set B, under Option 2 (variable Set B), it is more realistic for the gNB to vary among semi-fixed Set B including a limited number of deterministic Set B patterns rather than varying over totally random beams in Set B.
  + Proposal 6: For BM-Case-1 and Case-2, for the selection of Set B, consider Option 1 (Set B is fixed across training and inference) as a starting point.
    - For Option 2 (Set B is variable), semi-fixed Set B can be assumed in the evaluation, which includes a limited number of deterministic Set B patterns.
* Spreadtrum [4]:
  + Proposal 4: Set B to be a subset of set A for spatial domain beam prediction can be used as baseline,
    - If AI/ML inference is at NW side, beams in Set B can be determined by NW implementation.
    - If AI/ML inference is at UE side, beams in Set B can be determined with a fix pattern.
  + Proposal 6: For temporal beam prediction, evaluate and further study “Set A and Set B are the same” as the baseline.
* Vivo [5]:
  + (Option 1): only one pre-defined subset with fixed pattern in Set B is used in option 1 across training and inference stage,
  + while option 2 brings much more selection schemes in Set B, such as,
    - one fixed subset for training and another fixed subset for inference,
    - variable subsets with random patterns in Set B for training and inference, and
    - variable subsets with semi-random patterns in Set B for training and inference.
  + One fixed set B in Option 1 may show good performance in theory, but it lacks flexibility as in practical implementation, a particular beam or beam pair may suffer performance loss due to unexpected channel variation like blockage, and may cause large interference. Hence it is needed to study option 2 and make sure it can provide comparable performance as option 1 with higher flexibility.
  + ***Fixed beams:***
  + Observation 1: Fixed subset selection scheme with different fixed patterns brings tremendous performance difference.
  + Observation 2: Better performance gain can be obtained for one fixed subset selected by well-designed rule or enumerated with predefined searching criterion.
  + Observation 3: The performance with different training and validation fixed subsets is quite poor and not acceptable, i.e., fixed set B selection scheme suffers serious generalization issue.
  + Proposal 13: Unless an excellent generalization performance can be proved in option 1, i.e. a fixed subset in Set B for training and same fixed subset in Set B for validation, fixed set B selection scheme should be deprioritized.
  + ***Random subset selection***
  + Observation 4: Fixed beam subset in Set B can have good performance in ideal scenarios but it lacks flexibility. Issues like blockage and inter-cell interference can bring negative impact on the performance of fixed subset.
  + Observation 5: Random subset selection scheme, which allows multiple random subsets in training, can improve generalization performance as well as beam management related performance if compared to mismatched subset with always using one subset in training.
  + Observation 6: Set 5 with random beam subset still suffers tremendous performance deterioration due to huge number of combinations of selecting a target number of beams from total beam pairs.
  + Observation 7: Compared with Set 5, assistance information brings considerable gain in random subset selection scheme, especially for Tx/Rx beam angle as assistant information.
  + *Semi-random subset selection:*
  + Observation 8: To restrict the selection of random subset from the best X beam subsets can improve the performance of BM Case 1 prediction. Such semi-random selection with Tx/Rx beam angle information as input barely suffers performance loss compared with the best beam subset.
  + Observation 9: Semi-random beam subset scheme has potential to approach the performance upper bound, i.e. the best fixed subset, if the performance of each subset in top-N best subsets has similar performance of top-1 best subset.
  + Proposal 16: Support option 2 for Set B selected by semi-random beam subset selection scheme with both Tx and Rx beam information as AI input.
* OPPO [8]
  + Proposal 1: Apply fixed Set B across training and inference phases for BM-Case1 and BM-Case2.
* LGE [10]
  + Proposal 4. Option 1 can be considered as a baseline for selection of Set B of beams.
* CATT [12]
  + *Beam pair prediction:*
  + Observation 1: For Beam pair prediction, beam prediction accuracy with fixed pattern has better performance than random pattern, since beam ID is implicit in the fixed pattern.
  + Observation 2: For Beam pair prediction, additional Beam ID input have significant performance gain compared with beam prediction accuracy with random pattern using L1-RSRP input only.
  + *DL Tx beam prediction:*
  + Observation 3: For DL Tx beam prediction, beam prediction accuracy with fixed pattern has better performance than random pattern, since beam ID is implicit in the fixed pattern.
  + Observation 4: For DL Tx beam prediction, additional Beam ID input have significant performance gain compared with beam prediction accuracy with random pattern using L1-RSRP input only.
* Fujitsu [13]
  + Proposal 1: For selection of Set B of beams (pairs), it is suggested that the beams of fixed Set B are constructed with an even-spacing sampling rate from beams of Set A.
  + Proposal 2: For selection of Set B of beams (pairs), it is suggested that the beams of variable Set B are constructed with different even-spacing sampling rates from beams of Set A.
* Intel [14]
  + Proposal 2: The variability of Set B can only be due to updating the L1 measurements corresponding to beams or beam-pairs in Set B at different intervals. The cardinality of the set should not change across training and inference.
* Lenovo [15]
  + Allow set B to have variable number of beams at each instant of time during training and/or inference and allow the beams in set B to be variable and change across time during training and/or inference.
* Xiaomi [17]
  + Proposal 2: Adopt the evaluation methodologies listed below for spatial domain beam prediction:
    - Set B is a subset of set A.
    - AI model:
      * Input:
        + Scheme 2: L1-RSRP of beam pairs selected randomly and corresponding beam pair IDs
        + Scheme 3: L1-RSRP of beam pairs with fixed beam pair IDs
      * Output
        + L1-RSRP of all beam pairs with ascending order of beam pair ID
* Nokia [19]
  + Proposal 4: For BM-Case1, RAN1 further study Set B to be a fixed pattern.
  + Proposal 5: For BM-Case1 model inference applies at the NW side, with DL Tx beams considered for Set A and Set B, the training a model with random Set B is not needed.
  + Observation 8: For BM-Case1\_(DL Tx) model inference in UE side, training model with random Set B may reduce model switching/indication/ transferring overhead for UE. But the benefit of BM-Case1\_(DL Tx) model inference on the UE side is not yet clear.
  + Proposal 6: For BM-Case1 model inference applies at the UE side, with DL Tx beams considered for Set A and Set B, the training a model with random Set B can be further studied.
  + Observation 9: For Set B is different to Set A with Set B is wide beam, the KPI for the wide beam codebook design should be both prediction accuracy and throughput performance.
  + Proposal 7: For BM-Case1, RAN1 may further study the case of Set A/B are DL Tx and Set B/Set A are different.
    - Set B is a wide beam codebook and Set A is a refined beam codebook
    - Advance Set B designs are needed to provide sufficient refined beam prediction performance.
  + Observation: For DL Tx-Rx beam pair prediction, the prediction target space . And it requires a large number of measurements to have good beam pair prediction.
  + Proposal 8: RAN1 further investigates the comparison between independent Tx beam, Rx beam prediction, and joint Tx-Rx beam pair prediction.
  + Proposal 13: For BM-Case2, support Method 3 as the main scenario while other methods can be further studied :
    - Method 1: Set B is a fixed subset of Set A
    - Method 2: Set B is a variable subset of Set A
    - Method 3: Set B is the same as Set A
    - Methods 1 or 3 + Assistance Info: ML model input consists of L1-RSRP measurement based on Set B and assistance information
* MediaTek [20]:
  + Observation 9: With a greater number of beams in Set B, both models achieve higher Top-k accuracy. However, greater number of beams in Set B requires more beam RSRP measurements.
  + Proposal 8: Study the tradeoff between the beam measurement overhead and prediction accuracy for different number of beams in Set B.
  + Observation 10: The selection of beams in Set B will affect the prediction accuracy of the AI/ML-based spatial beam prediction.
  + Proposal 11: For AI/ML-based spatial domain beam prediction evaluation, study the subset selection (number and combination) if Set B is variable (Option2 on the selection of Set B of beams in the RAN1 #110 agreement).
  + Observation 11: The spatial beam prediction by using multi-arm beam design in Set B performs better than using subset beam design in Set B.
  + Observation 12: The spatial beam prediction by using wide beam design in Set B does not outperforms the performance by using subset beam design in Set B.
  + Proposal 12: Study and evaluate a more comprehensive Set B design, including joint designing the number of beams in Set B and their beam shape for spatial beam prediction.
* Samsung [24]
  + *DL Tx beam*
  + Observation # 1: Using the L1-RSRP of the “best” Rx beam with exhaustive beam sweep as inputs can provide the best performance for the accuracy of Top-1/N beam prediction than fixed or randomly selected one or two Rx beams with fixed or random Tx beams for BM-Case 1.
  + Observation # 2: With L1-RSRPs of fixed Rx beam(s) as AI inputs can provide better performance than L1-RSRP of random Rx beam(s) for DL Tx beam prediction for BM-Case 1.
  + Observation # 3: With L1-RSRP of fixed Tx beams in Set B of beams as AI inputs can provide better performance than with random Tx beam in Set B of beams for DL Tx beam prediction for BM-Case 1.
  + Observation # 7: For DL Tx beam prediction in BM-Case 1, L1-RSRPs with implicit Tx beam index as AI inputs and best Tx beam as AI outputs and can provide a better performance than with L1-RSRPs with implicit Tx beam index and Rx beam index as AI inputs and best Tx-Rx beam pair as AI outputs.
  + *Beam pair*
  + Observation # 4: Using the L1-RSRP of the “best” Rx beam with exhaustive beam sweep as inputs can provide the best performance for the accuracy of Top-1/N beam prediction than fixed one Rx beam or randomly selected one or two Rx beams with fixed or random Tx beams for DL Tx-Rx beam pair prediction for BM-Case 1.
  + Observation # 5: With L1-RSRPs of fixed Rx beam(s) as AI inputs can provide better performance than L1-RSRP of random Rx beam(s) for DL Tx-Rx beam pair prediction for BM-Case 1.
  + Observation # 6: For beam pair prediction for BM-Case 1, AI with inputs as L1-RSRPs of fixed Tx beams and implicit beam ID information can provide better performance than non-AI based approach.
* CEWiT [27]
  + Observation 1: When Set B is fixed, i.e., when fixed beam pattern is used, spatial domain beam prediction achieves sufficiently high performance with 50% overhead reduction.
  + Observation 2: When the size of Set B is increased, the performance of the AI/ML model improves.

Summary on the views on the different options for Set B of beams(pairs)

* Fixed beams
  + Huawei, Spreadtrum (as baseline), OPPO, LGE(baseline), Intel, xiaomi, Nokia(network side BM-Case 1 and BM-Case 2)
  + Concerns on generalization by vivo.
* Random beams
  + Lenovo(?), xiaomi
  + Intel (not preferred. only be due to updating the L1 measurements)
* Pre-configured beam patterns
  + Futurewei, Huawei, vivo, Nokia (UE side only), Nokia(network side BM-Case 2)

#### FL3: Set B of beams (Pairs)

**Please provide your views on the following questions:**

**Q1:** Whether fixed Set B of beams (Pairs) lacks of flexibility and may suffer from performance loss?

* Vivo: One fixed set B in Option 1 may show good performance in theory, but it lacks flexibility as in practical implementation, a particular beam or beam pair may suffer performance loss due to unexpected channel variation like blockage, and may cause large interference. Hence it is needed to study option 2 and make sure it can provide comparable performance as option 1 with higher flexibility.

**Q2:** Whether to support fixed beams across training and inference can be the baseline for BM-Case 1 and inference at gNB side?

**Q3:** For option 2, which of options do you support and which cases it needed, e.g., inference at UE side, Temporal domain prediction?

Opt A: Set B is variable with a pre-configured pattern in each time instant (e.g., for BM-Case 2) for each training

Opt B: Set B is randomly changed among pre-configured patterns (with fixed or variable number of beams(pairs)) in each report/measurement during training and/or inference

Opt C: Set B is randomly changed among Set A beams (pairs) (with fixed or variable number of beams(pairs)) in each report/measurement during training and/or inference

**D:** other comments

|  |  |
| --- | --- |
| Company | Comments |
| Google | Q1: No  Q2: Maybe yes. But we think more study is needed with regard to some model generalization related aspects.  Q3: It seems Opt C is more closed to legacy beam report |
| Xiaomi | Q1: Fixed Set B may have low generalization capability. But fixed set B can provide high beam prediction accuracy.  Q2: Both fixed set B and variable set B can be the baseline.  Q3: For temporal domain beam prediction, we prefer set B is same as set A for UE side inference. |
| OPPO | Q1: No  Q2: Support. We see no strong motivation to change beams during training and inference phases.  Q3: For BM-Case2, we prefer to adopt the same Set B and Set A. Since Set A is unchanged, so does Set B. |
| Spreadtrum | Q1: Maybe the generalization performance of fixed set B is not as good as that of random set B, but from the simulation results, fixed set B does have better performance of prediction accuracy.  Q2: YES.  Q3: For better generalization performance, option C is a more flexible choice. |
| NTT DOCOMO | Q1: No  Q2: Yes. The generalization evaluations are to verify the performance when the different configurations are used for training and inference. |
| Ericsson | We think Q2 is what can be agreed at this stage. |
| vivo | Q1: Yes.  Q2: No. We think even for gNB side inference, this will restrict gNB’s utilization of beam management in real network. Both fixed beams and variable beams need to be studied.  Q3: Both Opt A/B/C can be evaluated and studied. Based on our evaluation, Opt B with assistant information as input (e.g.., beam ID or angle) can achieve the best performance and high flexibility. However, we are not sure whether it is needed to make a decision on this at this stage. |
| FL1 | Please continue the discussion. |
| HW/HiSi | Q1: No  Q2: Yes  Q3: What is meant with Option 2 in this context? Could this please be clarified? For BM-Case 2, we support that Set B is a subset of Set A. It seems that Option A and Option B can be looked into further.  Agree with Ericsson that Q2 is what can be agreed at this stage. |
| Futurewei | Q1: the performance loss part may be true if the trained model (with fixed beam pattern) is used as inference when the input beam pattern is very different from the original pattern.  Q2: Yes |
| Qualcomm | Q1: No  Q2: Yes  To answer Q3, need to have evaluation results and look at the corresponding performance. Do not see the urgency to agree at this point. |
| Apple | The answers should be checked against generalization performance. It is easy to say Yes to Q1, but to say NO to Q1 needs lots of justification.  Q2 is related to generalization, e.g., infra vendor 1’s antenna configuration/analog beam design may be different from infra vendor 2’s, the design target dictates the answer to Q2. We feel it is premature to provide answer to any of the question. |
| CEWiT | Q1: No  Q2: Yes |
| CAICT | Q1: No  Q2: Yes  Q3: Not sure yet. |
| Samsung | Q1: No. At least in our understanding, when Set B beams is well distributed in Set A beams, so that it can provide sufficient information to obtain the best beam, even some of beam may be blocked. If vivo considered the case that gNB configure Set B beams close to each other, then it may suffer some blocking issue. We suggest to have some discussion on when/which procedure AI/ML for BM is aimed for to facility the discussion. It may also benefit to draw some observations and identify spec impact in later phase.  Q2: Yes. We think it can be baseline. We think for a gNB, the pattern can be controlled. Whether it may or may not have generalization issue, we can have separate discussions.  Q3: It is good to have some clarification. Listing or even agreeing on options may help for companies to clarify their assumption. However, we don’t see the need for down selection in this meeting. |
| Nokia | Q1: No  Q2: Yes.  Q3: It seems Opt C is more closed to legacy beam report |
| LG | Q1: No  Q2: Yes |
| CMCC | Q1: It restricts the flexibility in practical.  Q2: We prefer variable set B can be the baseline.  Q3: We support both Opt B and Opt C. |
| Fujitsu | Q1: No  Q2: Yes. But it needs to consider the generalization issues on the number of beams and their patterns in Set B. |
| ZTE | Q1: Fixed beam set B can obtain better performance than that of random beam set B. Nevertheless, it is still possible to have multiple fixed beam sets with different beam patterns or different numbers of beams for flexibility improvement and better balancing overhead and performance.  Q2: Both fixed beam set B and variable beam set B can be considered.  Q3: Both options can be further evaluated, which relates to the trade-off between performance and flexibility. |
| FL2 | |  | | --- | | **Agreement in RAN 1 #110**   * **Study the following options on the selection of Set B of beams (pairs)**    + **Option 1: Set B is fixed across training and inference**     - **FFS on the beams of Set B**   + **Option 2: Set B is variable (e.g., different beams (pairs) patterns in each report/measurement during training and/or inference)**      - **FFS on fixed or variable number of beams (pairs)**     - **FFS on the details**   + **Other options are not precluded.**   + **FFS on the number of beams (pairs) in Set B**   + **Note: This does not preclude the alternative that Set B is different from Set A.** |   Based on current inputs, and the agreements in RAN 1 #110, please share your view on the following proposal:  **Proposal 4-3-1a**   * **At least for BM-Case 1 with inference at gNB side, Set B is fixed across training and inference is baseline for evaluation.** * **FFS on Set B is variable**   Please continue to share your view on Q3. In FL’s view, it will be good to list options on the table for Option 2: Set B is variable, considering the options listed in Q3. |
| Lenovo | We do NOT support proposal 4-3-1a as it supports fixed Set B. Please read our reasoning and our responses to Q1, Q2, Q3.  Q1: Yes. Having fixed set of beams for measurement could lack the flexibility and may result in lower performance in some situations. For example, there could be a particular channel/network realization where the optimal beam corresponds to an uncommon/non-typical beam and the set of beams other than those fixed beforehand in Set B might give us more useful information in predicting the optimal beam.  FL: I think this depends on how to select/design the Set B. In my understanding, Set B is just a kind of compressed sensing, or down sampling of Set A. Since the information we need to know/predict is just one best beam, even fixed Set B shall give enough information, especially if it is well designed. Even some of measurements of some beams in Set B has small value, it provides information. If all measurements is poor, maybe the cell selection has some issues other than beam prediction.  Within enough data for training, random beam can provide acceptable performance, or may be even better performance (not observed by any of results yet).  Q2: No. Whether the AI/ML model is at UE or at gNB, considering only fixed beams is very restrictive as it would not allow using some of the powerful ML techniques (such as reinforcement/sequential learning).  FL: No intention to only study on fixed beam, please check the updated proposal.  Q3: Strongly support **Option C** for all cases of beam prediction. Option A and B are special cases of Option C. Thus, by adopting Option C, we will NOT eliminate Option A and B. Adopting either Opt A or Opt B, would prove to very restrictive and will not allow the companies the to explore all possible ML methods.  **Other Comments in support of variable Set B**:   1. Option 2 corresponds to a more generic way of selecting set B. Option 1 is a special case of option 2, where the cardinality (i.e., size) of set B is constant and set B contains the same beams all the time (in each measurement/reporting instance and during training and inference). Let denote the set B at time and let denotes the cardinality of set . In option 2, need not be equal to for and in option 1, for all times . 2. While an AI/ML model based on supervised learning/training can work with option 1, an AI/ML model based on online learning method may certainly require option 2. For example, in reinforcement learning, what beam to be probed in the next time instant depends on what beams have been measured so far (i.e., the actions taken so far) and what are their RSRP values (i.e., what are observed rewards). Thus, the beams to be measured are required to be selected on the fly, in a dynamic nature, making such methods difficult to work with option 1. Further, in such ML methods, the size/cardinality of set B as well as the beams (pairs) in set B vary for each instance of beam prediction.   *Please also refer to* **Section 4 in** [**R1-2209122**](https://www.3gpp.org/ftp/TSG_RAN/WG1_RL1/TSGR1_110b-e/Docs/R1-2209122.zip) *for more detailed arguments supporting variable Set B*.  Hence, we want Set B may be allowed to have a variable number of beams and elements of set B may be allowed to change at each instant of time. |
| CATT | We support the Proposal 4-3-1a.  Q1: No.  Q2: Yes.  Q3: can be further studied. |
| MediaTek | Q1: We think it can be beneficial to have variable Set B.  Q2: Considering with majority views, we support Proposal 4-3-1a.  Q3: Options A and B can be evaluated and studied, but we prefer to remove Option C as it is only limited to the case when Set B is a subset of Set A. |
| Futurewei | We prefer supporting both options, i.e., fixed patterns in Set B and variable patterns in Set B. Variable Set B patterns do not mean the patterns are random, they can be pre-determined. Companies should be given the flexibility of which option they would like to support.  We are ok with Proposal 4-3-1a but we think Option 2 for variable Set B should be included as specified in the original proposal while FFS can be on its details. |
| Intel | Q1: We think the answer is NO. But we also think companies are mixing the definition of what fixed Set B means. The beams provided in Set B can possibly change across training and inference if companies want to evaluate generalization performance across different samplings of Set A when Set B is a subset of Set A but the cardinality of set B does not need to change.  Q2: YES. We are ok in principle with Proposal 4-3-1a  Q3: We would like to clarify the intention behind the different options. Is this for testing generalization performance? Or is this for more robust training? We can consider Option B if the proponents can clarify how this helps with training.  FL: This is for basic scheme, not for generalization, which will be discussed separately. |
| CAICT | We can accept proposal 4-3-1a for performance comparison. |
| Apple | There may benefits in considering variable set B, e.g., in the generalization study. Settling on fixed set B simplifies the study, but the applicable condition for the resulted design may be too restrictive. |
| vivo | **Proposal 4-3-1a**  We think Set B with variable beams also needs to be evaluated for gNB side. If we only use one fixed set of beams, how could this study guide further system design and potential deployment? How can all cells using one fixed set to be measured by UE ensure good performance, considering some of the beams may suffer serious performance loss due to issues like blockage? Hence we think we need to consider and study both fixed and variable set B.  **Proposal 4-3-1a**   * **At least for BM-Case 1 with inference at gNB side, study both the case Set B is fixed and the case Set B is variable across training and inference ~~is baseline~~ for evaluation.** |
| ZTE | Q1: Fixed beam set B can obtain better performance than that of random beam set B. Nevertheless, it is still possible to have multiple fixed beam sets with different/pre-determined beam patterns or different numbers of beams for flexibility improvement and better balancing overhead and performance.  Q2: As agreed in RAN1#110, both fixed beam set B and variable beam set B can be considered. vivo’s update is fine to us.  Q3: Both options can be further evaluated, which relates to the trade-off between performance and flexibility. Besides, we prefer Opt A and Opt B since too random beam set B in Opt C may not provide performance guarantee. |
| FL2 | The original intention is not to drop variable Set B. Please consider proposal 4-3-1b.  Some further clarification on Options in Q3, this is listing the “potential options” for basic performance, not for generalization. In this meeting, I’d like to focus on provide a full list for further study with potential down selection if possible.  **Proposal 4-3-1b**   * **At least for BM-Case 1 with inference at gNB side, Set B is fixed across training and inference is baseline for evaluation.** * **~~FFS on~~ Set B is variable is optional, FFS on details** |
| Samsung | We support the proposal 4-3-1b. We believe Option 1 should be a baseline since it can be easily implemented by cell-specific DL-RSs for BM. Besides, we are open to discuss about benefit of Option 2 compared to Option 1.  Regarding Options in Q3, we think Opt B and Opt C may or may be operated randomly, so we suggest to remove ‘ randomly’ as follows:  Opt A: Set B is variable with a pre-configured pattern in each time instant (e.g., for BM-Case 2) for each training  Opt B: Set B is ~~randomly~~ changed among pre-configured patterns (with fixed or variable number of beams(pairs)) in each report/measurement during training and/or inference  Opt C: Set B is ~~randomly~~ changed among Set A beams (pairs) (with fixed or variable number of beams(pairs)) in each report/measurement during training and/or inference  FL3: In my understanding, if remove “randomly” in Opt B or Opt C, it will have overlap with Opt A. |
| Xiaomi | We prefer to take fixed set B and variable set B with same priority. |
| Ericsson | We support the proposal 4-3-1b. |
| CATT | We support the proposal 4-3-1b. |
| Qualcomm | Support 4-3-1b. |
| NTT DOCOMO | We could further study all options and companies could report their choice if they simulate the variable Set B. |
| Spreadtrum | Support proposal 4-3-1b. |
| HW/hiSi | We support the proposal 4-3-1b. |
| CMCC | Fine.  We think various number of Set B of beam(pairs) can be used for generalization performance verification. |
| FL3 | There are several companies support both options although slightly majority companies support to make fixed Set B as baseline. Two things will be helpful, one is to have some focus, the other is to list options on the table for companies to further study and evaluate.  Based on current situation, please consider the following proposal. Please check three options for Option 2.  **Proposal 4-3-1c**   * **Study the following options on the selection of Set B of beams (pairs)**    + **Option 1: Set B is fixed across training and inference**   + **Option 2: Set B is variable (e.g., different beams (pairs) patterns in each report/measurement during training and/or inference), FFS:**     - **Opt A: Set B is variable with a pre-configured pattern in each time instant (e.g., for BM-Case 2) for each training**     - **Opt B: Set B is randomly changed among pre-configured patterns (with fixed or variable number of beams(pairs)) in each report/measurement during training and/or inference**     - **Opt C: Set B is randomly changed among Set A beams (pairs) (with fixed or variable number of beams(pairs)) in each report/measurement during training and/or inference**   + **Other options are not precluded.** * **At least for BM-Case 1 with inference at gNB side, Option 1(Set B is fixed across training and inference) is prioritized for evaluation.** |
| NVIDIA | Support Proposal 4-3-1c |
| Futurewei | We can support Proposal 4-3-1c. We think Option B (specified in Option 2) is also applicable for BM-Case1. |
| InterDigital | We are fine with Proposal 4-3-1c in principle, but have one question.  For Opt B/C, Set B is changed during training and/or inference. However, for Opt A, Set B changes only for training. Is this intended? If so, then what would be the reason? If not, we prefer to add inference as well. |
| Xiaomi | First, for Option 2, Opt A, Opt B and Opt C are not at the same level since Opt A is talking about the set B in different time instance within one sample only for BM Case 2. But Opt B and Opt C are talking about the set B in different sample for both BM Case 1 and Case 2. We suggest to separate the discussion on BM Case 1 and Case 2, and propose the following update   * + **Option 2: Set B is variable (e.g., different beams (pairs) patterns in each report/measurement during training and/or inference), FFS:**     - **For BM case 1**       * **Opt A: Set B is randomly changed among pre-configured patterns (with fixed or variable number of beams(pairs)) in each report/measurement during training and/or inference**       * **Opt B: Set B is randomly changed among Set A beams (pairs) (with fixed or variable number of beams(pairs)) in each report/measurement during training and/or inference**     - **For BM Case 2**       * **Opt A: Set B is randomly changed among pre-configured patterns (with fixed or variable number of beams(pairs)) in each report/measurement during training and/or inference**         + **Opt A-1: Set B is variable with a pre-configured pattern in different measurement time instance within one sample**         + **Opt A-2: Set B is fixed in different measurement time instance within one sample**       * **Opt B: Set B is randomly changed among Set A beams (pairs) (with fixed or variable number of beams(pairs)) in each report/measurement during training and/or inference**         + **Opt B-1: Set B is variable with a pre-configured pattern in different measurement time instance within one sample**         + **Opt B-2: Set B is variable among Set A beams (pairs) (with fixed or variable number of beams(pairs)) in different measurement time instance within one sample**         + **Opt B-3: Set B is fixed in different measurement time instance within one sample**     Second, for the “**At least for BM-Case 1 with inference at gNB side, Option 1(Set B is fixed across training and inference) is prioritized for evaluation.**” We think if for evaluation only, there is no difference between gNB-side inference and UE-side inference. But the difference is the spec impact and the generalization performance. If generalization performance will be not considered in this proposal, we are fine to consider fixed set B first for BM Case 1 with inference at gNB side. |
| ZTE | In the present description, the difference between Opt A and Opt B is not clear. We may consider merge Opt A and Opt B into one option to avoid confusing.   * + **Option 2: Set B is variable (e.g., different beams (pairs) patterns in each report/measurement during training and/or inference), FFS:**     - **Opt A/B: Set B is ~~randomly~~ changed among pre-configured patterns (with fixed or variable number of beams(pairs)) with/without a pre-determined order** **in each report/measurement during training and/or inference**     - **Opt C: Set B is randomly changed among Set A beams (pairs) (with fixed or variable number of beams(pairs)) in each report/measurement during training and/or inference**   Besides, the last bullet can be revised as   * **At least for BM-Case 1 ~~with inference at gNB side~~, Option 1(Set B is fixed across training and inference) is prioritized for evaluation.** |
| CMCC | One clarification question on Opt C.  According to the wording, the meaning of Opt C is that set B can be randomly changed among Set A with randomly or pre-configured patterns. If set B is changed among pre-configured patterns, it will be both Opt B and Opt C. |
| OPPO | Support the proposal. |
| HW/HiSi | We are fine in principle with the proposal Proposal 4-3-1c, for the last bullet, we share ZTE’s view. |
| CATT | To CMCC:  We think the Opt C is set B can be randomly changed among Set A without pre-configured patterns, which is different with Opt B. If our understanding is correctly, we prefer to add “without pre-configured patterns” in Opt C.  For the last bullet, we also share ZTE’s view. |
| Samsung | Regarding the second bullet, we believe FL’s intention is Option 1 is baseline at least for BM-Case1 in DL Tx beam prediction.  Regarding the Opt A, we think this may be used for BM-Case 1 as well as inference but not sure this time. We just want to hear other companies’ view about Opt A. Therefore, we have following modification:    **Proposal 4-3-1c**   * **Study the following options on the selection of Set B of beams (pairs)**    + **Option 1: Set B is fixed across training and inference**   + **Option 2: Set B is variable (e.g., different beams (pairs) patterns in each report/measurement during training and/or inference), FFS:**     - **Opt A: Set B is variable with a pre-configured pattern in each time instant [(e.g., for BM-Case 2)] for each training [and/or inference]**     - **Opt B: Set B is randomly changed among pre-configured patterns (with fixed or variable number of beams(pairs)) in each report/measurement during training and/or inference**     - **Opt C: Set B is randomly changed among Set A beams (pairs) (with fixed or variable number of beams(pairs)) in each report/measurement during training and/or inference**   + **Other options are not precluded.** * **At least for BM-Case 1 with inference for DL Tx beam prediction ~~at gNB side~~, Option 1 (Set B is fixed across training and inference) is prioritized for evaluation.** |
| vivo | We don’t agree with the last bullet. These are different solutions rather than simulation assumptions for alignment among companies. We think using a fixed set of set B beams is not practical at all in real deployment. Hence we don’t think we should prioritize this specific solution.   * **~~At least for BM-Case 1 with inference at gNB side, Option 1(Set B is fixed across training and inference) is prioritized for evaluation.~~** |
| LG | For the last bullet, we support ZTE’s version. |
| Spreadtrum | For the last bullet, we support ZTE’s version. |
| Lenovo | Thanks for revising the proposal. We support 4-3-1c.  However, we think it is not good and not at all desirable to prioritize fixed Set B over variable Set B as it would restrict the type of AI/ML techniques that can be used for BM in the study. The objective of the study is to see the potential of AI/ML methods and we should not curtail the freedom to explore, and employ, different AI/ML techniques. We would like the last bullet to be removed.  **~~At least for BM-Case 1 with inference at gNB side, Option 1(Set B is fixed across training and inference) is prioritized for evaluation.~~** |
| Ericsson | Agree with ZTE’s version on last bullet |
| MediaTek | We prefer to remove Option C as it is only limited to the case when Set B is a subset of Set A.  Same view with InterDigital, we wonder why Opt A is just for training, we think OptA can be applied to model inference as well. Also, we suggest to remove “**(e.g., for BM-Case 2)”** from OptA, we like the example but it seems like a lot of companies may think OptA is only for BM Case 2. However, in our view, OptA is applicable to BM Case1 as well, if assistance information such as beam ID is provided to the model. Therefore, we suggest the following changes:   * + **Option 2: Set B is variable (e.g., different beams (pairs) patterns in each report/measurement during training and/or inference), FFS:**     - **Opt A: Set B is variable with a pre-configured pattern in each time instant ~~(e.g., for BM-Case 2)~~ for each training and/or inference**     - **Opt B: Set B is randomly changed among pre-configured patterns (with fixed or variable number of beams(pairs)) in each report/measurement during training and/or inference**     - **~~Opt C: Set B is randomly changed among Set A beams (pairs) (with fixed or variable number of beams(pairs)) in each report/measurement during training and/or inference~~**   We also prefer to remove the last bullet for “**At least for BM-Case 1 with inference at gNB side, Option 1(Set B is fixed across training and inference) is prioritized for evaluation.**”. It is too early to prioritize one option at this stage. |

## 4.4 Assumption of time domain information for BM-Case 2

There were some discussions/proposals/disclosures of assumption on the beam sweeping and RS pattern:

* Huawei/HiSi [2]
  + Proposal 19: Considering the robustness of AI/ML against longer prediction time instances and higher UE speeds, CSI-RS patterns for the observation/prediction window should be assumed with a large time domain distance between observation/prediction instances, for example 80ms or 160ms.
* vivo [5]
  + BM evaluation metrics are calculated based on difference between decided/predicted best beam pair and real best beam pair in T2. For comparison, non-AI and AI based 2-step scheme are evaluated. For AI based 2-step scheme, best pair is predicted based on P2+P3 procedure, and for non-AI 2-step scheme, best pair is decided based on measurement in P2+P3 procedure without prediction. Time duration T1 is fixed to 8\*40ms, and time duration T2 is equal to 1\*40ms, 4\*40ms or 8\*40ms respectively.
* Ericsson [11]
  + The NN’s inputs at training and inference are the L1-RSRPs selected from 5 consecutive time instances. So, the observation duration T1=5\*40ms=200ms. Prediction is at the time instance immediately following the last observation window time instance, and also a prediction at 160ms ahead for comparison. Hence the time duration for the best beam evaluation is T2= 40 ms or 160 ms.
* Xiaomi [17]
  + Observation 5: AI based beam prediction scheme 1 and scheme 2 in time domain can provide good performance.
    - Scheme 1 assumes same periodicity for history measurement instance and future time instance.
    - Scheme 2 assumes longer periodicity for history measurement instance than that of future time instance. It can reduce more RS overhead than scheme 1.
  + Proposal:
    - Set A and set B are the same set.
    - The periodicity of future time instance can be 80ms/160ms
    - The periodicity of future time instance can be same or different from that of history measurement instance
    - AI model:
      * Input:
        + L1-RSRP of set B in 4 history measurement instances
      * Output
        + Top K beams of set A in 1/2/4 future instances
* Nokia [19]:
  + CSI measurement/report periodicity: 40ms or 80ms
  + Observation window: 200, Prediction window: 40 80ms
* Mediatek [20]:
  + Proposal 3: Evaluate the impact of different observation and prediction window sizes to the performance of AI/ML temporal beam prediction.
  + Observation 2: By fixing the observation window size, the accuracy performance becomes better when prediction window size is lower.
  + Observation 3: By fixing the prediction window size, the accuracy performance increases when the observation window size increases. However, the performance will saturate.
  + Proposal 5: When the prediction window size is fixed, evaluate and study the optimal observation window size in terms of prediction accuracy and RS overhead.
* DoCoMo [25]
  + Proposal 3: Both Pattern 1 and Pattern 2 should be considered in the BM-Case 2 for both evaluation and study on specification impact.
    - Pattern 1: The sequence of inputs of AI/ML model has different periodicity and time scale from that of the sequence of outputs (Input: large; Output: small).
    - Pattern 2: The sequence of inputs of AI/ML model has the same periodicity and time scale as that of the sequence of outputs.
* 
* **Figure 3.** Illustration of pattern 1
* 
* **Figure 4.** Illustration of pattern 2

#### FL3: Assumptions for BM-Case 2

**Proposal 4-4-1a:**

* **At least for BM-Case 2, consider the following assumptions for evaluation**
  + **Periodicity of time instance for each measurement/report:**
    - **[20ms], 40ms, 80ms, 160ms**
    - **Other values can be reported by companies.**
  + **Number of time instances for measurement/report:** 
    - **4, [5], 8**
    - **Other values can be reported by companies.**
  + **Time instance(s) for prediction:**
    - **[20ms], 40ms, 80ms, 160ms, [1440ms] after the last [time instance/measurement/report]**
    - **Other values can be reported by companies.**

**Please provide your view Proposal 4-4-1a, if any.**

|  |  |
| --- | --- |
| Company | Comments |
| FL1 | * The intention is to align the assumption for evaluation results collection. * The numbers proposed/used by single company are put in bracket. * The definition of time instance(s) for prediction needs to be discussed, e.g., after the last time instance, or measurement or report. * In FL’s view, with separated periodicity for measurement and prediction, DoCoMo’s proposal can be covered. |
| Google | OK with the proposal in principle. But it seems 20ms periodicity is a good point since it is a typical SSB periodicity. |
| Xiaomi | First we need to clarify that the periodicity for measurement/report depends on the number of Rx beam for measurement. It means that the periodicity for measurement/report equals to N times of the periodicity of SSB/CSI-RS, where N is the number of Rx beam for measurement.  In addition, as for the 3rd sub-bullet, does it mean the ‘time duration for prediction’? We prefer the similar description as measurement time instance.  And we suggest following update  **Proposal 4-4-1a:**   * **At least for BM-Case 2, consider the following assumptions for evaluation**   + **Periodicity of time instance for each measurement/report:**     - **[20ms], 40ms, 80ms, 160ms**     - **Other values can be reported by companies.**   + **Number of time instances for measurement/report:**      - **4, [5], 8**     - **Other values can be reported by companies.**   + **Periodicity of time instance(s) for prediction:**     - **[20ms], 40ms, 80ms, 160ms,**     - **Note: the periodicity of time instances for prediction can be same or smaller than that of time instance for each measurement/report**     - **Other values can be reported by companies.**   + **Number of time instances for prediction:**      - **1, 2, 4, 8**     - **Other values can be reported by companies.**   FL1: In my understanding, whether there is one or more prediction values as AI output (time instance(s) for prediction) is companies’ choice. I don’t want to give impression that we require to have multiple output by one AI model, which is different from inputs. That’s the reason I didn’t use the same structure for measurement and prediction.  However, companies are encouraged to share the views. |
| NTT DOCOMO | We wonder whether the last bullet intends ‘the periodicity of time instance(s) for prediction’ or ‘the longest duration for the time instance(s) for prediction’. If it is the previous one, we think ‘the number of time instances for prediction’ is also needed.  FL1: Please find my comments above. |
| vivo | OK |
| FL1 | Please continue the discussion. |
| HW/HiSi | For the number of instances of measurement and report, the candidates are 4, [5] and 8. Is this the number of instances for combined observation and prediction, or only for one of them? If it is the latter, we think the numbers are too large. |
| NVIDIA | Support |
| Qualcomm | We suggest removing the second and fourth bullet. The size of the input sequence as well as the output sequence for predictions does not need to be specified, but it can be reported by the companies. |
| Xiaomi | @FL, thanks for your response. So your intention is to consider only one time instance? If yes, we think our proposed update is much clearer and can include the number of time instance is 1. If you have concern on more than one time instance, we can update it to [2, 4, 8] and hear more comments.  **Proposal 4-4-1a:**   * **At least for BM-Case 2, consider the following assumptions for evaluation**   + **Periodicity of time instance for each measurement/report:**     - **[20ms], 40ms, 80ms, 160ms**     - **Other values can be reported by companies.**   + **Number of time instances for measurement/report:**      - **4, [5], 8**     - **Other values can be reported by companies.**   + **Periodicity of time instance(s) for prediction:**     - **[20ms], 40ms, 80ms, 160ms,**     - **Note: the periodicity of time instances for prediction can be same or smaller than that of time instance for each measurement/report**     - **Other values can be reported by companies.**   + **Number of time instances for prediction:**      - **1, [2, 4, 8]**     - **Other values can be reported by companies.**   FL1: In original proposal there is “(s)” after “time instance”, it means that one or more time instance can be reported by companies. No need to define periodicity. |
| CAICT | Support. |
| OPPO | For BM-Case2, mobility models including Option 2 and 3 apply the granularity of 100ms.  **Option #2: Linear trajectory model with random direction change.**  **Option #3: Linear trajectory model with random and smooth direction change.**   * + - UE moving trajectory: UE will move straightly along the selected direction to the end of a~~n~~ time interval, where the length of the time interval is provided by using an exponential distribution with average interval length, e.g., 5s, with granularity of 100 ms.   With above being said, can we also suggest to add 100ms as one candidate of periodicity of time instances for measurement/report and prediction.   * + **Periodicity of time instance for each measurement/report:**     - **[20ms], 40ms, 80ms, [100ms], 160ms**     - **Other values can be reported by companies.**   + **Time instance(s) for prediction:**     - **[20ms], 40ms, 80ms, [100ms], 160ms, [1440ms] after the last ~~[~~time instance ~~/~~for measurement/report~~]~~**     - **Other values can be reported by companies.** |
| Nokia | Ok with the proposal. |
| LG | OK |
| ZTE | We prefer xiaomi's update. Besides, it has been agreed in agenda 9.2.3.2 that the number of time instances for measurement and prediction is up to implementation and thus doesn't need to be specified. Otherwise, we suggest to add more candidates for the number of time instances for prediction.   * + **Number of time instances for prediction:**      - **1, 2, 4, 5, 8**     - **Other values can be reported by companies.** |
| FL2 | Please provide your view in the following proposal.  **Proposal 4-4-1b:**   * **At least for BM-Case 2, consider the following assumptions for evaluation**   + **Periodicity of time instance for each measurement/report in T1:**     - **[20ms], 40ms, 80ms, [100ms],160ms**     - **Other values can be reported by companies.**   + **Number of time instances for measurement/report in T1:**      - **4, [5], 8**     - **Other values can be reported by companies.**   + **Time instance(s) for prediction in T2:**     - **[20ms], 40ms, 80ms, [100ms], 160ms, [1440ms] after the last [time instance/ measurement/report]**     - **Other values can be reported by companies.**     - **One or more time instance can be reported by companies.** |
| CATT | Fine with Proposal 4-4-1b. |
| MediaTek | OK with the latest proposal |
| Intel | Ok with latest proposal. We think 20ms measurement periodicity should be included. |
| CAICT | Support proposal 4-41b. |
| vivo | OK with the updated proposal from FL |
| ZTE | Fine with the latest proposal. Besides, we suggest to add 400ms and 800ms to the candidates of time instances for prediction. |
| Samsung | Generally fine with the FL proposal. We suggest to remove blanket of [20ms] in the first sub-bullet which is typical value for SSB burst sharing the same view with Intel. |
| Xiaomi | We still have some concern on the 3rd sub-bullet. The 3rd sub-bullet is talking about the time instance(s), so the value listed in the sub-sub-bullet is the value for one time instance or multiple instances or for T2? And what is the relationship between time instance(s) and T2? In my understanding, T2 can be divided into multiple time instances with same duration.  Thus we still prefer the revision in our previous comments  **Proposal 4-4-1b:**   * **At least for BM-Case 2, consider the following assumptions for evaluation**   + **Periodicity of time instance for each measurement/report in T1:**     - **[20ms], 40ms, 80ms, [100ms],160ms**     - **Other values can be reported by companies.**   + **Number of time instances for measurement/report in T1:**      - **4, [5], 8**     - **Other values can be reported by companies.**   + **~~Time instance(s) for prediction in T2:~~**     - **~~[20ms], 40ms, 80ms, 160ms, [1440ms] after the last [time instance/ measurement/report]~~**     - **~~Other values can be reported by companies.~~**     - **~~One or more time instance can be reported by companies.~~**   + **Periodicity of time instance(s) for prediction:**     - **[20ms], 40ms, 80ms, [100ms], 160ms,**     - **Note: the periodicity of time instances for prediction can be same or smaller than that of time instance for each measurement/report**     - **Other values can be reported by companies.**   + **Number of time instances for prediction:**      - **1, [2, 4, 8]**     - **Other values can be reported by companies.**   FL3: Please check whether the update for prediction can resolve your concern or not. |
| Ericsson | Ok with FL update. Also propose to remove the blanket of [20ms], it is a commonly used value. |
| Qualcomm | We still believe the second bullet should be removed and we do not need to specify the number of time instances used for measurement as input to the AI/ML model, as it is an implementation choice and could be left up to companies and companies can report the assumptions. For the third bullet, we suggest not mentioning the absolute values of possible time instances. This does not well capture the possible options that we can have for measurement/prediction instances. The terminology that we have adopted in our contribution is as follows:  FL3: the purpose is just for collaboration results and understand the performance of BM-Case2 Companies need to report the number of inputs.  “An MxNy beam prediction formulation is used: in this formulation, the prediction algorithm is given as input L1-RSRP measurements from x contiguous beam management cycles out of every x+y contiguous cycles, then provides predictions for the best L1 beams for the following y cycles.”  FL3: I think current number of time instances is your x. and periodicity of time instance on beam management cycle.  For instance, we may be measuring every 100ms, and performing predictions for the next 20ms,40ms, 60ms, 80ms, etc. Or we may be measuring every 160ms, and performing predictions for the next 40ms, 80ms, 120ms, etc. This is just an illustrative example that shows why third bullet is not inclusive of the possible options. Of course, we can add 60ms, 120ms, etc. but then we would be exhaustively listing many options which questions the existence of the third bullet in the first place. Also propose to remove brackets for 20ms. |
| LG | Ok with FL update. |
| NTT DOCOMO | Since the ‘**Time instance(s) for prediction in T2**’ does not mean the periodicity of the predicted results while the possible time duration after the last [time instance/ measurement/report] instead, it is not clear for us how to decide the time durations between the last [time instance/ measurement/report] and the predicted values if more than 1 time instances are predicted.  Therefore, we support Xiaomi to modify the 3rd bullet with following refinement:   * + **Periodicity of time instance(s) for prediction in T2:**     - **[20ms], 40ms, 80ms, [100ms], 160ms,**     - **Note: the periodicity of time instances for prediction can be same or smaller than that of time instance for each measurement/report**   **Other values can be reported by companies.**  FL3: please check the updated proposal |
| HW/HiSi | We think the instances in T1 should also include 2. The reason is that the more instances we use for observation, the more complex the AI model becomes, and also the beam sweeping overhead might increase. It will be useful to also evaluate the performance of less complex schemes.    **Updated Proposal 4-4-1b:**   * **At least for BM-Case 2, consider the following assumptions for evaluation**   + **Periodicity of time instance for each measurement/report in T1:**     - **[20ms], 40ms, 80ms, [100ms],160ms**     - **Other values can be reported by companies.**   + **Number of time instances for measurement/report in T1:**      - **2, 4, [5], 8**     - **Other values can be reported by companies.**   + **Time instance(s) for prediction in T2:**     - **[20ms], 40ms, 80ms, [100ms], 160ms, [1440ms] after the last [time instance/ measurement/report]**     - **Other values can be reported by companies.**     - **One or more time instance can be reported by companies.** |
| FL3: | Tried merge Xiaomi’s proposal int new one. I am hesitate to using “periodicity” for prediction, because in my understanding, one time instance shall be the baseline although I understand that companies also look into more instances. Please consider the following proposal:  **Proposal 4-4-1c:**   * **At least for BM-Case 2, consider the following assumptions for evaluation**   + **Periodicity of time instance for each measurement/report in T1:**     - **~~[~~20ms~~]~~, 40ms, 80ms, [100ms], 160ms**     - **Other values can be reported by companies.**   + **Number of time instances for measurement/report in T1:**      - **[2], 4, [5], 8**     - **Other values can be reported by companies.**   + **Time duration ~~instance(s)~~ for prediction ~~in T2~~:**     - **~~[~~20ms~~]~~, 40ms, 80ms, [100ms], 160ms, [1440ms] after the last [time instance/ measurement/report]**     - **Other values can be reported by companies.**     - **One or more time instance can be reported by companies in the time duration.** |
| InterDigital | We do not prefer adding additional evaluation assumption over and over and prefer to focus on the approved evaluation assumptions.  However, we can accept this proposal as a compromise if the proposed evaluation assumptions are optional. |
| Xiaomi | Support the Proposal 4-4-1c |
| ZTE | We are not sure whether one time instance for prediction can be a baseline. According to simulation results provided by some companies, beams of multiple future time instances can be predicted well. Besides, according to our preliminary simulation results, the AI based method can achieve satisfactory gains over the non-AI method with a time duration of 400/800ms for prediction. Thus, we suggest to add 400ms and 800ms to the candidates of time duration for prediction.   * + **Time duration ~~instance(s)~~ for prediction ~~in T2~~:**     - **~~[~~20ms~~]~~, 40ms, 80ms, [100ms], 160ms, [400ms], [800ms], [1440ms] after the last [time instance/ measurement/report]** |
| OPPO | Support in principle.  For the last bullet “Time duration for prediction”, we understand it is as per instance, rather than F prediction duration. If that’s the case, we suggest to try the following wording   * + **Time duration ~~instance(s)~~ for prediction per instance ~~in T2~~:** |
| HW/HiSi | Removing the brackets from [2] for the number of time instances in T1 would be better, but we can live with it for progress. |
| CATT | Fine with Proposal 4-4-1c. |
| Samsung | If we change from “time instance” to “time duration”, the together with “after the last [time instance/ measurement/report]” is unclear. For example, according to definition of time duration, the time duration is from 0ms to the next predicted time instance, so it becomes BM-Case1. We prefer the original wording of proposal 4-4-1b. |
| LG | Support the Proposal 4-4-1c |
| NTT DOCOMO | We are fine with proposal 4-4-1c. |
| Lenovo | Support |
| Ericsson | Support Proposal 4-4-1c |
| MediaTek | Support |

## 4.5 (on hold) Assistance information

Assistance information were discussed and some observations are summarized:

* Huawei [2]
  + Observation 3: For the AI/ML-based beam prediction, the provision of some assistance information may be infeasible due to the concern of disclosing proprietary information or privacy to the other side. For a NW-side model, this includes Rx beam angle or boresight direction, Rx beam shape, and FFS the UE speed and UE position. For a UE-side model, a list of infeasible assistance information includes at least the Tx beam angle or boresight direction, 3dB beamwidth, and Tx beam shape.
    - The meaning and method to obtain expected Tx/Rx beam information, LOS probability may need to be clarified before discussing.
* ZTE[3]
  + We also note that other assistance information such as beam shape or beam usage are not evaluated since they are implementation-related information of the gNB or UE, which may be difficult to be disclosed to the opposite node
* Vivo [5]:
  + Observation 7: Compared with Set 5, assistance information brings considerable gain in random subset selection scheme, especially for Tx/Rx beam angle as assistant information.
  + Proposal 14: Assistance information, such as Tx/Rx beam ID or angle in connection with input RSRPs, should be used as AI input with random subset selection for both BM-Case1 and BM-Case2.
  + Proposal 15: Suggest to use both Tx and Rx beam information as assistance information for further performance improvement in random subset selection.
  + Proposal 23: Further study assistance information, such as beam shape pattern, 3dB beam width, etc., as model input to address performance deterioration for generalization of different beam shapes in BM-Case1.
  + ***BM-Case 2: with different beam shape patterns***
  + Observation 18: For the case using local beam ID as model input, beam loss and accuracy degenerate significantly compared to the performance of AI model training and inference with beam pointing angle.
  + Proposal 27: Further study assistance information, such as beam shape pattern, 3dB beam width, etc., as model input to address performance deterioration for generalization of different beam shapes in BM-Case2.
  + Proposal 28: Suggest to use beam pointing angle or global beam ID as assistance information for AI model input.
* OPPO [8]
  + Proposal 4: For the assistance information of BM-Case1 and BM-Case2, suggest to
    - Justify the performance benefits if assistance information
    - Study whether assistance information would expose beamforming implementation and proprietary information at NW or UE.
* Nokia [19]
  + Observation 4: For BM-Case1, the ML model using as input only RSRP measurements has performances that reduce significantly changing the number of RSRP measurements from 8 to 4, i.e. further down sampling Set A, from a ratio of ¼ to a ratio of 1/8.
  + Observation 5: For BM-Case1, when the ML model use the UE angle as the assistance information, it has a better performance than all the other variants.
  + Observation 6: For BM-Case1, the ML model using as input RSRP measurements and UE Position has performances that outweigh the performance of the ML model using only RSRP.
  + Observation 7: For BM-Case1, using assistance information like Beam Angle and Beam ID related to the measured beams may not significantly improve the performance of the ML model using as input only RSRP with a fixed pattern.
  + Proposal 3: For BM-Case1, RAN1 further study the use of assistance information at the ML model input. The following assistance information can be prioritized:
    - the beam angle and/or the beam boresight direction for the measured DL Tx beams from NW to UE.
    - the UE position information.
    - the UE’s angle relative to a panel array of the gNB
* MediaTek [20]:
  + Observation 6: Temporal beam prediction by adding additional UE angle information directly to the input of the model did not show significant gains compared to predicting without UE angle information.
  + Proposal 6: Study more scenarios where additional information may improve the temporal beam prediction performance
  + Observation 13: The spatial prediction accuracy does not improve much by using UE angles directly as the additional input, at least for the ratio of Set B and Set A sizes is between 1/6 to 1.
  + Observation 14: The spatial prediction accuracy does not improve much by using UE angles directly as the additional input, under various selections of Set B.

## 4.6 Others

Some other input/output related discussion:

* Futurewei [1]
  + Observation 1: The performance of spatial-domain beam prediction using either fixed beam pattern sampling or pre-configured beam patterns sampling is improving when the training dataset size is increasing.
  + Observation 3: Performance of pre-configured beam patterns sampling is more sensitive to training dataset size increase compared to fixed beam pattern sampling.
* Vivo [5]:
  + Proposal 5: At least AI model inputs/outputs and training/validation dataset should be reported per sub-use case by companies. Other parameters, such as NN architecture type, loss function, and data post/pre-processing method, are encouraged to be reported.
* Ericsson [11]
  + To help enable reproducibility, companies should report relevant information about the AI/ML model architecture, data pre- and post-processing, loss functions, and training procedures using an academic style paper and/or pseudocode.

|  |
| --- |
| **Agreement**   * Companies are encouraged to report the following aspects of AI/ML model in RAN 1 #110. FFS on whether some of aspects need be defined or reported.   + Description of AI/ML model, e.g, NN architecture type   + Model inputs/outputs (per sub-use case)   + Training methodology, e.g.     - Loss function/optimization function     - Training/ validity /testing dataset:       * Dataset size, number of training/ validity /test samples       * Model validity area: e.g., whether model is trained for single sector or multiple sectors       * Details on Model monitoring and model update, if applicable   + Others related aspects are not precluded |

#### FL2: (close)Other assumptions

**Q: Any other assumption needs to be aligned/reported for evaluation results collection? E.g., amount of training/testing data set?**

|  |  |
| --- | --- |
| Company | Comments |
| NTT DOCOMO | We would like to encourage other companies to report which RS (SSB or CSI-RS) is assumed in the simulation. In case of SSB, Rx beam sweeping should be considered to obtain all the Tx-Rx beam pairs of SetB, while CSI-RS can ignore it by using repetition.  FL1: I think this is a valid point, especially for BM-Case2. Maybe some clarification/discussion on the assumption of RS for each time instance after agreed on Proposal 4-4-1a. |
| NTT DOCOMO | We should study the RS overhead reduction/RS overhead for different cases of BM-Case 2 separately, i.e. same or different periodicities within T1 and T2. |
| FL3 | Let’s close the discussion here |
|  |  |

# Evaluation results for AI/ML in beam management

#### FL3: Results collection

**Proposal 5-1a: => Proposal 5-1b (with updates) => Proposal 5-1c (with updates)**

For both BM-Case1 and BM-Case 2, the following table is adopted as **working assumption** for reporting the evaluation results.

**Table X. Evaluation results for AI/ML model deployed on [UE or network]-side without model generalization for [DL Tx beam prediction or Tx-Rx beam pair prediction or Rx prediction] [with Y beams [pair] in Set A]**

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| AI/ML model Input/output | | Data size | | AI/ML model | | | Evaluation results | | | | | |
| [Beam prediction accuracy (%)] | | [L1-RSRP Diff] | [System performance] | | |
| Model input | Model output | Training | testing | [Short model description] | ~~Model~~ Inference complexity | Computational complexity | KPI A | KPI B… | [Average L1-RSRP diff] | [RS overhead] | [UCI report] | [UTP]  … |
|  |  |  |  |  |  |  |  |  |  |  |  |  |

To report the following in table caption:

* + - * Which side the model is deployed

Further info for the columns:

* Model input: input type, e.g., L1-RSRP and the number of beams in Set B
* Model output: output type, e.g., the best DL Txand/or Rxbeam ID, ~~and the number of beams in Set A~~
* Dataset size, both the size of training/validation dataset and the size of test dataset
* Short model description: e.g., CNN, LSTM
* ~~AI/ML~~ Inference complexity:
  + ~~both~~ model ~~inference~~ complexity in terms of “number of model parameters” and/or size (e.g. Mbyte)”, and
  + computational complexity in terms of FLOPs
* Evaluation results: agreed KPIs

Note: To report other simulation assumptions, if any.

**Please provide your views on the template for result collection**

|  |  |
| --- | --- |
| Company | Comments |
| Google | Shall we add number of beams in set A? I guess number of beams in set B can be covered by data size.  FL1: updated as part of model input/output. Also, welcome for other suggestion, e.g. if we can align the number, different tables can be created for different number of beams in Set A can be created. |
| FL1: | Updated the proposals |
| vivo | Issues like set B beam construction (fixed or variable), # of output beams, utilization of assistant information, set A-set B relationship, beam pair or beam prediction, etc., are included in model input/output?  FL1: Added in the title of the Table. |
| Nokia | Set B/A dimensions and type may be needed to report. Update looks ok. |
| FL2 | Please comment on the proposals |
| Lenovo | It is desirable to clarify the difference between “Model Inference Complexity” and “Computational Complexity” mentioned in the table. |
| CATT | 1. The number of beams in Set A is not related with model output. Thus, suggest to add it in the title of the Table or have a separate column to describe Set A and Set B.  2. For model output, we think the intent is output the number of best DL Tx and/or Rx beam ID. Also suggest to add “L1-RSRP” as model output. Thus suggest the following update for model output:  Model output: output type, e.g., the number of best DL Tx and/or Rx beam ID and the corresponding L1-RSRP~~number of beams in Set A~~  3. For AI/ML complexity, we have same concern with Lenovo that what means “Model Inference Complexity” in the table. Suggest to align with proposal “both model complexity in terms of “number of model parameters”, and computational complexity in terms of FLOPs”. Thus change “Model Inference Complexity” into “Model Complexity” in the table, which is in terms of “number of model parameters”. |
| MediaTek | We have same question as Lenovo, what is the difference between “Model inference complexity” and “Computational complexity”? We wonder if there is an agreed definition for the first term. |
| Futurewei | We are ok with the proposal in general, but we agree with CATT that “Model Inference Complexity” should be change to “Model Complexity”. Another point is that beam prediction accuracy may be different in different scenarios, thus, we suggest at least adding scenario/configuration as another attribute in the table.  Another point is that we need to have another report template for companies to report model generalization evaluation results. Please consider the attributes/template for model generalization results provided in our contribution for BM EVM (R1-2208368).   |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | | **Sub use case** | **Training scenario / config.** | **Testing scenario/ config.** | **Set A/B configurations** | **Dataset size** | | **Perf. KPIs** | **Other KPIs** | **Mechanism applied** | | Train | Test | |  |  |  |  |  |  |  |  |  | |
| Intel | OK in general. Model Inference Complexity should be changed to Model Complexity as commented by others. |
| CAICT | Fine with the proposal to use Model complexity instead of model inference complexity. |
| vivo | Generally ok with the proposal from FL. |
| FL2 | @Lenovo, CATT, MTK, please refer to the agreements made in 9.2.1  @Futurewei, It will be considered later after this and generalization proposal.   * *Inference complexity*   + *Computational complexity of model inference: FLOPs*   + *Computational complexity for pre- and post-processing*   + *Model complexity: e.g., the number of parameters and/or size (e.g. Mbyte)*   It is updated in **Proposal 5-1b (with updates),** on which, please share your view. |
| Samsung | We support the proposal 5-1b |
| Ericsson | Support 5-1b. |
| CATT2 | After double checking, update our comments as following:  1. For DL Tx beam prediction, the type of model output maybe DL Tx beam ID. But for DL beam pair or Rx beam prediction, the type of model output should also include DL Rx beam ID. So we suggest to add DL Rx beam ID in the proposal 5-1b. Moreover, the corresponding L1-RSRP can be also as a model output since one agreed KPI is average L1-RSRP difference.  2. The number of beams in Set A is NOT a kind of model output. Why do we add “the number of beams in Set A” in the type of model output? Suggest description “the number of beams in Set A” in the title of the Table or in a separate column.  3. For model inference complexity, thanks FL for pointing out the agreement in 9.2.1. After double check with the agreement, it’s agreed that the model inference complexity includes computational complexity (i.e., FLOPs) and model complexity (e.g., number of parameters and/or size). Thus, we think it’s better to align with the agreement in 9.2.1.  Based on above comments, we suggest following updates on Proposal 5-1b.  **Table X. Evaluation results for AI/ML model deployed on [UE or network]-side without model generalization for [DL Tx beam prediction or Tx-Rx beam pair prediction or Rx prediction] with number of beams in Set A**   |  |  |  |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | | AI/ML model Input/output | | Data size | | AI/ML model | | | Evaluation results | | | | | | | [Beam prediction accuracy (%)] | | [L1-RSRP Diff] | [System performance] | | | | Model input | Model output | Training | testing | [Short model description] | Model ~~inference~~ complexity | Computational complexity | KPI A | KPI B… | [Average L1-RSRP diff] | [RS overhead] | [UCI report] | [UTP]  … | |  |  |  |  |  |  |  |  |  |  |  |  |  |   To report the following in table caption:   * + - * Which side the model is deployed   Further info for the columns:   * Model input: input type, e.g., L1-RSRP and the number of beams in Set B * Model output: output type, e.g., the best DL Tx and/or Rx beam ID and the corresponding L1-RSRP ~~the number of beams in Set A~~ * Dataset size, both the size of training/validation dataset and the size of test dataset * Short model description: e.g., CNN, LSTM * ~~AI/ML~~ Inference complexity:   + model ~~inference~~ complexity in terms of “number of model parameters and/or size (e.g. Mbyte)”, and   + computational complexity in terms of FLOPs * Evaluation results: agreed KPIs   Note: To report other simulation assumptions, if any.  FL3: Most suggestions are considered. There is no discussion/KPIs for L1-RSRP and model output. |
| Qualcomm | There is an ongoing discussion about the “common KPIs” across use cases in 9.2.1. Suggest referring the AI/ML model complexity definition to the outcome of the discussions in 9.2.1 to avoid duplicate discussions. The metrics mentioned here are under the umbrella of “common KPIs” and not specific to beam prediction use cases. |
| LG | Agree with Qualcomm. |
| NTT DOCOMO | Generally fine with the proposal. |
| Spreadtrum | We support the proposal 5-1b |
| FL3 | Agreements in 9.2.1 in RAN 1 #110   * *Inference complexity*   + *Computational complexity of model inference: FLOPs*   + *Computational complexity for pre- and post-processing*   + *Model complexity: e.g., the number of parameters and/or size (e.g. Mbyte)*   The current proposal already followed agreements in 9.2.1. The reason to propose this as WA other than agreement is to be able to update based on any new agreements in 9.2.1 or 9.2.3.1.  Most of KPIs are in bracket, which is also for the purpose of updates.  Please share your view on **Proposal 5-1c (with updates)** |
| Futurewei | As FL clarified, “Inference complexity” includes computational complexity (for model inference and pre- and post-processing) and model complexity, what does the other “Computational complexity” mean in the table (in 5-1c)? |
| Xiaomi | Suggest the following update on model output  Model output: output type, e.g., the best DL Txand/or Rxbeam ID, and/or L1-RSRPs |
| ZTE | Generally fine with the proposal. |
| CATT | Thanks FL for explanation and considering our suggestions. We still have the following comment.  In the update proposal, inference complexity includes computational complexity and model complexity. But in the table, there are still inference complexity and computational complexity. We think they are not paratactic concepts. Suggest to change inference complexity into model complexity as our previous comments.   |  |  |  |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | | AI/ML model Input/output | | Data size | | AI/ML model | | | Evaluation results | | | | | | | [Beam prediction accuracy (%)] | | [L1-RSRP Diff] | [System performance] | | | | Model input | Model output | Training | testing | [Short model description] | Model ~~inference~~ complexity | Computational complexity | KPI A | KPI B… | [Average L1-RSRP diff] | [RS overhead] | [UCI report] | [UTP]  … | |  |  |  |  |  |  |  |  |  |  |  |  |  | |
| Lenovo | Support proposal 5-1C. However, when “Inference complexity” includes ‘computational complexity of model inference’ and ‘computational complexity for pre and post processing’, it is not clear what should be reported in the “Computational complexity” column in table in Proposal 5-1C. |
| MediaTek | Agree with Futurewei and CATT, if we want to follow agreements in 9.2.1, we think it is better to use “Model complexity” instead of “Inference complexity” to align with the level of “Computational complexity” in the agreement. |

## (on hold) Evaluation results for BM-Case 1

The following observations were provided in contributions:

* Futurewei [1]:
  + Observation 2: The performance of AI/ML-based beam prediction has achieved significantly better performance when comparing with sparse beam sweeping approach.
* Huawei/HiSi [2]:
  + Observation 4: For spatial domain beam prediction, AI/ML-based schemes under the 64-DFT codebook outperform the legacy approach in most of the cases in terms of beam selection accuracy, e.g.,:
    - AI/ML-based Top-5 prediction reaches almost the upper performance bound with a prediction accuracy of 94.95% but with an overhead reduction of 67.17%. On the other hand, for the same overhead reduction, the established legacy Baseline approach can only achieve a prediction accuracy of 55.3%
    - With AI/ML-based Top-3 prediction, the overhead compared to the legacy Baseline approach can be further reduced by another 8%, while the prediction still is much higher (89.2% as opposed to 55.3%)
  + Observation 5: For spatial domain beam prediction, AI/ML-based schemes under the 64-DFT codebook outperform the legacy approach in most of the cases in terms in terms of average L1-RSRP difference, e.g.,:
    - For AI/ML-based Top-5 prediction, the L1-RSRP difference compared to genie-aided beam prediction in Exhaustive 64 is as low as 0.03 dB, with an overhead reduction of 67.17%. On the other hand, for the same overhead reduction, the established legacy Baseline approach can only achieve an average L1-RSRP difference of 1.02dB
    - With AI/ML-based Top-3 prediction, the overhead compared to the legacy Baseline approach can be further reduced by another 8%, while the average L1-RSRP difference is still is much smaller (0.08dB as opposed to 1.02dB)
  + Observation 6: It can be observed that better prediction accuracy is achieved when Set B is a subset of Set A compared to the case where Set B is a wide beam set, especially when K=1; with the increase of K, the gap between two options becomes narrower.
* ZTE [3]:
  + Proposal 3: The AI/ML model can be utilized for spatial domain beam prediction, which can greatly reduce the RS overhead for measurement while maintain a high beam prediction accuracy.
* Interdigital [6]
  + Observation 7: AI aided beam selection achieves more than 95% selection accuracy when error margin is larger than 0.5 dB by consuming 50%/33% of the measurement overhead for the exhaustive measurement.
  + Observation 8: AIML-based RSRP estimation always outperforms the baseline especially when less RSRP measurements are available as it achieves a higher selection accuracy by 35% when error margin is 0.5 dB.
* China Telecom [7]
  + Observation 1: Modelling the spatial beam prediction task as a classification model provides better performance with less training overhead.
* OPPO [8]
  + Spatial domain beam prediction can yield beam prediction accuracy (at least 80%) while overhead/latency reduction rate is 75%.
  + The system level metric, i.e. spectrum efficiency or throughput, is not sensitive to the L1-RSRP difference introduced by spatial domain beam prediction.
  + For 80% of the incorrect spatial domain beam prediction cases, the L1-RSRP difference can be kept within 2dB.
  + When beam prediction accuracy is high (at least 80%) and L1-RSRP difference is small (within 1 dB), the system-level performance, i.e. spectrum efficiency or throughput, may only provide non-essential insight, therefore focusing on L1-RSRP for beam prediction would be good enough.
* Ericsson [11]
  + *Tx beam prediction (with RSRP from best Rx beam)*
  + Observation 3: In outdoor scenarios, AI/ML can reduce beam spatial-domain beam prediction overhead substantially while maintaining good accuracy, both for 4x8 (30 beams in Set A) and 8x16 arrays (168 beams in Set A).
  + Observation 4: In scenarios with primarily indoor UEs, spatial-domain beam predication is more challenging.
  + *Tx/Rx beam prediction*
  + Observation 5: Joint TX/RX prediction can give good performance while significantly reducing RS overhead compared to measurements of all RX beams for each TX beam in Set B.
  + *System level performance*
  + Observation 6: The gains from AI/ML over baseline algorithm in terms of basic KPIs translate well to gains in full-buffer system-level evaluations.
  + *Reporting overhead*
  + Observation 7: By allowing variable number of reported beams via UE pre-processing of measurements, the reporting overhead can be substantially reduced with little performance degradation.
    - We consider a scheme with gNB-side inference where the UE measures a fixed set of beams, but only reports beams with RSRP exceeding a certain threshold relative to the strongest beam, i.e. only beams with an RSRP at most X dB below the RSPR of the strongest measured beam are reported.
* CAICT [16]
  + Observation 1: AI-based solution could achieve good performance for beam pair prediction with same training and validation set configurations.
* CMCC [18]
  + Observation 1: The increase of K significantly improves the prediction accuracy while leading to a small degree of increased beam sweeping overhead.
  + Observation 2: Compared with baseline option 1, AI based spatial beam prediction has minor loss of prediction accuracy for top-K beam pair but has large beam sweeping overhead reduction.
  + Observation 3: Compared with baseline option 2, AI based spatial beam prediction significantly enhances prediction accuracy for top-K beam pair under the same beam sweeping overhead.
* Rakuten Symphony [22]
  + Observation 1: The probability of one of the K beams being the best beam is more than 95% for K = 4.
  + Proposal 1: Consider a two-step beam management procedure where legacy beam management mechanism is used to choose the best beam from a set of beam recommendations from the AI/ML model.
* NVDIA [22]:
  + Observation 2: AI/ML-based algorithms for beam prediction in spatial domain can achieve performance comparable to that of exhaustive beam search, while the reference signal overhead, measurement effort, reporting overhead, and latency can be much reduced.
* Samsung [24]:
  + Observation # 10: For spatial domain prediction, AI can provide better performance in terms of beam prediction accuracy than non-AI based scheme with the measurements of a given subset of beams to select a best beam among a full set of beams.
  + Observation # 11: With the help of AI, SSB/RS overhead for measurements, UE measurement efforts, reporting overheads can be reduced to achieve a target performance for beam selection.
  + Observation # 12: For spatial domain prediction, AI can provide better performance in terms of beam prediction accuracy than non-AI based scheme with the measurements of a set of wide beams and a subset of narrow beams to select a best beam among a full set of narrow beams.
  + Observation # 13: For spatial domain prediction, AI can predict the best narrow beam based on the measurements of wide beams only with decent performance.
  + Observation # 14: For spatial domain prediction, AI can help gNB to predict the best narrow beam set that including the best narrow beam for UE to measure with high probability.

## (on hold) Evaluation results for BM-Case2

The following observations were provided in contributions:

* Huawei/HiSi [2]:
  + Observation 7: The AI/ML-based beam prediction based on the Set A with 256 beams (Type-2) provides a considerable gain over the legacy upper bound Exhaustive 64 (Type-1) in achievable L1-RSRP for a small fraction of the overhead associated with an Exhaustive 64 sweep.
  + Proposal 17: For AI/ML-based temporal domain beam prediction, regarding the relationship between Set A and Set B:
    - The size of Set B smaller than Set A should be considered as baseline.
      * Both can be considered in evaluations: Set B is a subset of Set A; Set B contains wide beams with full direction which are different from Set A with narrow beams.
    - Set B equal to Set A can be optionally used for performance comparison in evaluations.
  + Observation 8: For temporal beam prediction, AI/ML based methods are more robust than legacy approaches to variations of the UE speed.
    - When the time instance is 0.08s in the observation and prediction window, for UE speed 30km/h, the AI/ML Top-8 approach is 42% better than for the legacy baseline but for a UE speed of 90 km/h, the AI/ML Top-8 prediction accuracy is 47% better than for the legacy baseline
    - When the time interval is 0.16s in the observation and prediction window, for UE speed 30km/h, the AI/ML Top-8 approach is 48% better than for the legacy baseline but for UE speed 90 km/h, the AI/ML Top-8 prediction accuracy is 77% better than for the legacy baseline.
  + Observation 9: For temporal beam prediction, lower spatial consistency has more impact on the prediction accuracy achieved by the legacy approach than on accuracy achieved by the AI/ML-based methods. This can be seen from the results when different time instances are evaluated.
    - For UE at 30km/h, the accuracy of AI/ML Top-8 degrades 3.35% but the baseline degrades 4.8% when stretching the two prediction instances from 0.08s to 0.16s
    - For UE at 90km/h, the accuracy of AI/Ml Top-8 degrades 0.93% but the baseline degrades 9.56% when stretching the two prediction instances from 0.08s to 0.16s
* ZTE [3]:
  + Observation 5: A better beam prediction accuracy is achieved if more measured RSRPs are input to the AI model. However, for a NW-side model, more measured RSRPs used as AI input also means increasing of the UE reporting overhead.
  + Observation 6: Compared with the selected non-AI method, a more significant performance gain is observed if the beam set for measurement is a subset of the beam set for prediction.
* Vivo [5]:
  + Observation 15: For BM-Case2, compared with non-AI scheme, beam pair prediction scheme improves beam prediction accuracy and reduces average L1-RSRP difference significantly.
  + Proposal 24: Further study beam pair prediction scheme with expected information as AI input for improving generalization performance in BM-Case2.
* Interdigital [6]
  + Observation 9: AIML-based beam selection achieves more than 95% selection accuracy when error margin is larger than 0.5 dB by consuming 50% of the measurement overhead of the exhaustive measurement and it also shows better accuracy when the error margin is low.
  + Proposal 14: Further study benefits of AI/ML aided beam prediction.
* OPPO[8]
  + Temporal domain beam prediction can provide beam prediction accuracy (at least 77%) while overhead/latency reduction can be up to 50% (for the case of K = 4 and F = 4).
  + Beam predication accuracy slightly decreases from 87.1% to 77.1% (the case of Top-1) when F increases from 1 to 4, but strongly increases from 77.1% to 98.8% (the case of F = 4) when predicted beam number increases from Top-1 to Top-4.
  + For 80% of the incorrect temporal domain beam prediction cases, the L1-RSRP difference is lower than 3.5dB which may not strongly impact the spectrum efficiency.
  + Spatial and temporal domain beam prediction can provide beam prediction accuracy (at least 74.4%) while overhead/latency reduction can be up to 87.5% (for the case of K = 4, F = 4 and Set B = 32 beam pairs, Set A = 128 beam pairs).
  + Spatial and temporal domain beam prediction can provide beam prediction accuracy (at least 64.5%) while overhead/latency reduction can be up to 87.5% (for the case of K = 8, F = 8 and Set B = 32 beam pairs, Set A = 128 beam pairs).
* Ericsson [11]
  + Observation 10 The observed prediction performance improvement over baseline when number of beams in set B is <=8 is mainly due to the spatial domain prediction ability
  + Observation 11 With set A equal to set B and having 30 km/h straight line moving UEs with no rotation, AI/ML temporal prediction at T2=40ms shows no gain over baseline method due to the slow-varying channel.
* Xiaomi[17]
  + Observation 6: Set B < set A causes much more performance degradation compared to set B=set A for temporal beam prediction.
* Nokia[19]
  + Observation 18: For BM-Case2, the ML model using as input only RSRPs has performance that decreases when Set B is a subset of Set A and if no advanced algorithm is applied for beam selection in Set B.
  + Observation 19: For BM-Case2, the ML model using as input only RSRPs has performance that decreases when increasing the length of the prediction window.
  + Observation 20: For BM-Case2, additional algorithm (i.e. Bayesian Optimization) should be applied for choosing the beam measurements in Set B for the scenario of Set B is a subset of Set A.
  + Proposal 14: For BM-Case2, with Set B is a subset of Set A, measurement instances K and prediction instances F shall be carefully investigated prior supporting the sub-use case.
* Mediatek [20]:
  + *Performance between different models*
  + Observation 4: Transformer performs better than LSTM in terms of Top-k accuracy, and it requires less observation window size than LSTM does to achieve the same level of RSRP difference.
  + Observation 5: The computing complexity of Transformer is larger than LSTM, furthermore, the computing complexity increases with the observation window for both models.
  + Proposal 4: For different choices of prediction and observation window sizes, study the optimal model for to use, considering their computing complexity, UE’s computational and storage capacity.
  + Observation 7: Tx beam prediction’s Top-k performance is better than beam pair prediction’s Top-k performance. However, beam pair prediction doesn’t require UE Rx beam sweeping during the prediction windows.
  + Proposal 7: Study the tradeoff between using Tx beam prediction or beam pair prediction mechanisms considering their prediction Top-k accuracy, and corresponding beam management overhead.
* NVIDIA [23]
  + Observation 3: AI/ML-based algorithms for beam prediction in time domain can simply use a history of the best beam index to perform the prediction.
  + Observation 4: AI/ML-based algorithms for beam prediction in time domain can help lower reference signal overhead and reduce UE’s measurement requirement.
* Samsung [24]
  + Observation # 17: For time and spatial domain prediction, AI can provide better performance in terms of beam prediction accuracy than non-AI based scheme with the measurements of a subset of narrow beams to select a best beam among a full set of narrow beams.
* DoCoMo [25]
  + Observation 2: AI/ML could improve the beam prediction accuracy in time-domain, and the performance gain is higher in the high UE speed scenario.
  + Observation 3: The performance of AI/ML-based beam prediction is good even if Rx-sweeping periodicity (P) is large (>>20ms).
  + Proposal 4: Discuss the different performance gain for different UE speed for BM-Case 2, and consider the target scenario/speed for BM-Case 2 .

Observation 4: Similar tendency to pattern 1 could be observed for pattern 2 while the absolute performance gain of AI/ML is lower.

# Others

Some companies suggest to consider multiple scenarios for evaluations.

* Huawei/HiSi: [2]
  + Proposal 7: The evaluation for beam prediction should focus on a one-sided AI/ML model.
  + FL0: will be discussed in 9.2.3.2
* Interdigital [6]
  + Proposal 7: Support ‘Set B is a subset of Set A’ when Set A and Set B are utilized in a same frequency range for both temporal/spatial domain prediction.
  + Proposal 8: Support ‘Set A and Set B are different’ when Set A and Set B are utilized in different frequency ranges for both temporal/spatial domain prediction.
  + Proposal 9: AI/ML based beam management based on association between different frequency ranges should supported for both between FR1 and FR2-1 and between FR2-1 and FR2-2.
  + Proposal 10: For conventional scheme to obtain performance KPIs, current specification for beam management (i.e., up to 4 CRIs with L1-RSRP/SINR or SRS based prediction) should be considered.
  + Proposal 11: Number of beams in Set B should be decided and reported by each company.
  + FL0: Suggest to propose to in 9.2.3.2
* Intel [14]:
  + Proposal 1: For AI/ML evaluation for beam management use cases, including spatial and temporal domain beam management, consider only offline training of AI/ML models.
  + FL0: will be discussed in 9.2.3.2
* Nokia [19]
  + Observation 16: Selecting the beam based on the QoS based model output can improve the throughput performance of each UE by clustering the UEs to a single beam.
  + Observation 17: Even with errors in the input RSRP values of beams which are not in set B, QoS based beam selection can improve the throughput performance of each UE.
  + Proposal 11: For BM-Case1, RAN1 further investigate QoS-based beam prediction with predicted RSRPs of the beams in set A.
  + FL0: Suggest to propose to in 9.2.3.2
* Mediatek [20]:
  + Observation 8: Transformer always outperforms DNN in both datasets under various sizes of Set B. However, Transformer is more complex than DNN in terms of FLOPs.
  + Observation 15: For spatial beam prediction, the prediction performance of the best beam pair by using Transformer is only around 10% worse than predicting the best Tx beam.
  + Observation 16: For spatial beam prediction, the prediction performance of the best beam pair by using DNN is significantly worse than predicting the best Tx beam.
  + Proposal 13: Further study the use of larger size AI/ML model for best Tx/Rx beam pair prediction in spatial beam prediction.
  + FL 0: based on the conclusion in framework, this study does not intend to compare the performance of different AI models.

|  |
| --- |
| **Conclusion**  As indicated in SID, although specific AI/ML algorithms and models may be studied for evaluation purposes, AI/ML algorithms and models are implementation specific and are not expected to be specified. |

# Proposals for GTW on 10/10

**Proposal 3-1-1a: (Same agreements as in 9.2.2.1)**

**The following cases are considered for verifying the generalization performance of an AI/ML model over various scenarios/configurations as a starting point:**

* **Case 1: The AI/ML model is trained based on training dataset from one Scenario#A/Configuration#A, and then the AI/ML model performs inference/test on a dataset from the same Scenario#A/Configuration#A**
* **Case 2: The AI/ML model is trained based on training dataset from one Scenario#A/Configuration#A, and then the AI/ML model performs inference/test on a different dataset than Scenario#A/Configuration#A, e.g., Scenario#B/Configuration#B, Scenario#A/Configuration#B**
* **Case 3: The AI/ML model is trained based on training dataset constructed by mixing datasets from multiple scenarios/configurations including Scenario#A/Configuration#A and a different dataset than Scenario#A/Configuration#A, e.g., Scenario#B/Configuration#B, Scenario#A/Configuration#B, and then the AI/ML model performs inference/test on a dataset from a single Scenario/Configuration from the multiple scenarios/configurations, e.g., Scenario#A/Configuration#A, Scenario#B/Configuration#B, Scenario#A/Configuration#B.**
  + - * **Note: Companies to report the ratio for dataset mixing**
      * **Note: number of the multiple scenarios/configurations can be larger than two**
* **FFS the detailed set of scenarios/configurations**
* **FFS other cases for generalization verification, e.g.,**
  + - * **Case 2A: The AI/ML model is trained based on training dataset from one Scenario#A/Configuration#A, and then the AI/ML model is updated based on a fine-tuning dataset different than Scenario#A/Configuration#A, e.g., Scenario#B/Configuration#B, Scenario#A/Configuration#B. After that, the AI/ML model is tested on a different dataset than Scenario#A/Configuration#A, e.g., subject to Scenario#B/Configuration#B, Scenario#A/Configuration#B.**

**Proposal 1-1-2a:**

* **For system performance related KPI (if supported) [e.g, throughput] evaluation (model inference), companies report the traffic model:**
  + **Option 1: Full buffer**
  + **Option 2: FTP model with detail assumptions (e.g., FTP model 1, FTP model 3)**

**Proposal 1-1-1b:** (updated)

* **BS antenna configuration:** 
  + **antenna setup and port layouts at gNB: (4, 8, 2, 1, 1, 1, 1), (dV, dH) = (0.5, 0.5) λ**
  + **Other assumptions are not precluded**
* **BS Tx power:** 
  + **40dBm**
  + **Other values (e.g. 34 dBm) are not precluded and can be reported by companies**
* **UE antenna configuration (Clarification of agreement in RAN 1 #110):** 
  + **antenna setup and port layouts at UE: (1, 4, 2, 1, 2, 1, 1), 2 panels (left, right)**
  + **Other assumptions are not precluded**

**Proposal 2-1-1a:**

* **The definition of beam prediction accuracy (%) for Top-1 and/or Top-K beams:**
  + **Option 1 (optional): The beam prediction accuracy (%) is the percentage of “the Top-1 predicted beam is one of the Top-K genie-aided beams”**
  + **Option 2 (baseline): The beam prediction accuracy (%) is the percentage of “the Top-1 genie-aided beam is one of the Top-K predicted beams”**

**Proposal 4-2-1a:**

* **Adopt the following proposals as working assumption:**
* **For the evaluation of both BM-Case1 and BM-Case 2, 32 or 64 [or 256] Tx beams are used at NW side.** 
  + **Other values are not precluded and can be reported by companies.**
* **For the evaluation of both BM-Case1 and BM-Case 2, 4 Rx beams per UE panel are used at UE side.** 
  + **Other values are not precluded and can be reported by companies.**

# Reference

[1] R1-2208368, Continued discussion on evaluation of AI/ML for beam management, FUTUREWEI

[2] R1-2208431, Evaluation on AI/ML for beam management, Huawei, HiSilicon

[3] R1-2208523, Evaluation on AI for beam management, ZTE

[4] R1-2208549, Evaluation on AI for beam management, Spreadtrum Communications

[5] R1-2208636, Evaluation on AI/ML for beam management, vivo

[6] R1-2210240, Discussion for evaluation on AI/ML for beam management, InterDigital, Inc.

[7] R1-2208771, Evaluation on AI/ML for beam management, China Telecom

[8] R1-2208852, Evaluation methodology and preliminary results on AI/ML for beam management, OPPO

[9] R1-2208880, On Evaluation of AI/ML based Beam Management, Google

[10] R1-2208901, Evaluation on AI/ML for beam management, LG Electronics

[11] R1-2208906, Evaluation on AI/ML for beam management, Ericsson

[12] R1-2208969 Evaluation on AI/ML for beam management, CATT

[13] R1-2209013 Evaluation on AI/ML for beam management, Fujitsu

[14] R1-2209049 Evaluations for AI/ML beam management, Intel Corporation

[15] R1-2209122 Evaluation on AI/ML for beam management, Lenovo

[16] R1-2209232 Some discussions on evaluation on AI-ML for Beam management , CAICT

[17] R1-2209279 Evaluation on AI/ML for beam management, xiaomi

[18] R1-2209330 Discussion on evaluation on AI/ML for beam management, CMCC

[19] R1-2209369 Evaluation of ML for beam management, Nokia, Nokia Shanghai Bell

[20] R1-2209508 Evaluation on AI/ML for beam management, MediaTek Inc.

[21] R1-2209578 Evaluation on AI/ML for beam management, Apple

[22] R1-2209613 Evaluation of AI/ML based beam management, Rakuten Symphony

[23] R1-2209627 Evaluation of AI and ML for beam management, NVIDIA

[24] R1-2209724 Evaluation on AI ML for Beam management, Samsung

[25] R1-2209898 Discussion on evaluation on AI/ML for beam management, NTT DOCOMO, INC.

[26] R1-2209978 Evaluation on AI/ML for beam management, Qualcomm Incorporated

[27] R1-2210107 Evaluation on AI/ML for beam management, CEWiT

# Appendix: Agreements

# Agreements in RAN 1 #109e

[**R1-2205269**](file:///C:\Users\feifei.sun\AppData\Local\Temp\Docs\R1-2205269.zip) **Feature lead summary #1 evaluation of AI/ML for beam management Moderator (Samsung)**

From May 17th GTW session

Agreement

* For dataset construction and performance evaluation (if applicable) for the AI/ML in beam management, system level simulation approach is adopted as baseline
  + Link level simulation is optionally adopted

Agreement

* At least for temporal beam prediction, companies report the one of spatial consistency procedures:
  + Procedure A in TR38.901
  + Procedure B in TR38.901

Agreement

* At least for temporal beam prediction, Dense Urban (macro-layer only, TR 38.913) is the **basic** scenario for dataset generation and performance evaluation.
  + Other scenarios are not precluded.
* For spatial-domain beam prediction, Dense Urban (macro-layer only, TR 38.913) is the **basic** scenario for dataset generation and performance evaluation.
  + Other scenarios are not precluded.

Agreement

* At least for spatial-domain beam prediction in initial phase of the evaluation, UE trajectory model is not necessarily to be defined.

Agreement

* At least for temporal beam prediction in initial phase of the evaluation, UE trajectory model is defined. FFS on the details.

[R1-2205270](file:///C:\Users\feifei.sun\AppData\Roaming\Microsoft\Docs\R1-2205270.zip) Feature lead summary #2 evaluation of AI/ML for beam management Moderator (Samsung)

[R1-2205271](file:///C:\Users\feifei.sun\AppData\Roaming\Microsoft\Docs\R1-2205271.zip) Feature lead summary #3 evaluation of AI/ML for beam management Moderator (Samsung)

**Decision:** As per email decision posted on May 20th,

Agreement

* UE rotation speed is reported by companies.
  + Note: UE rotation speed = 0, i.e., no UE rotation, is not precluded.

Agreement

* For AI/ML in beam management evaluation, RAN1 does not attempt to define any common AI/ML model as a baseline.

Conclusion

Further study AI/ML model generalization in beam management evaluating the inference performance of beam prediction under multiple different scenarios/configurations.

* FFS on different scenarios/configurations
* Companies report the training approach, at least including the dataset assumption for training

Agreement

* For evaluation of AI/ML in BM, the KPI may include the model complexity and computational complexity.
  + FFS: the details of model complexity and computational complexity

Agreement

* For spatial-domain beam prediction, further study the following options as baseline performance
  + Option 1: Select the best beam within Set A of beams based on the measurement of all RS resources or all possible beams of beam Set A (exhaustive beam sweeping)
    - FFS CSI-RS/SSB as the RS resources
  + Option 2: Select the best beam within Set A of beams based on the measurement of RS resources from Set B of beams
    - FFS: Set B is a subset of Set A and/or Set A consists of narrow beams and Set B consists of wide beams
    - FFS: how conventional scheme to obtain performance KPIs
    - FFS: how to determine the subset of RS resources is reported by companies
  + Other options are not precluded.

**Decision:** As per email decision posted on May 22nd,

Agreement

* For dataset generation and performance evaluation for AI/ML in beam management, take the parameters (if applicable) in Table 1.2-1b for Dense Urban scenario for SLS

**Table 1.2-1b Assumptions for Dense Urban scenario for AI/ML in beam management**

|  |  |
| --- | --- |
| **Parameters** | **Values** |
| **Frequency Range** | FR2 @ 30 GHz   * SCS: 120 kHz |
| **Deployment** | 200m ISD,   * 2-tier model with wrap-around (7 sites, 3 sectors/cells per site)   Other deployment assumption is not precluded |
| **Channel mode** | UMa with distance-dependent LoS probability function defined in Table 7.4.2-1 in TR 38.901. |
| **System BW** | 80MHz |
| **UE Speed** | * For spatial domain beam prediction, 3km/h * For time domain beam prediction: 30km/h (baseline), 60km/h (optional) * Other values are not precluded |
| **UE distribution** | * FFS UEs per sector/cell for evaluation. More UEs per sector/cell for data generation is not precluded. * For spatial domain beam prediction: FFS:   + Option 1: 80% indoor ,20% outdoor as in TR 38.901   + Option 2: 100% outdoor * For time domain prediction: 100% outdoor |
| **Transmission Power** | Maximum Power and Maximum EIRP for base station and UE as given by corresponding scenario in 38.802 (Table A.2.1-1 and Table A.2.1-2) |
| **BS Antenna Configuration** | * [One panel: (M, N, P, Mg, Ng) = (4, 8, 2, 1, 1), (dV, dH) = (0.5, 0.5) λ as baseline] * [Four panels: (M, N, P, Mg, Ng) = (4, 8, 2, 2, 2), (dV, dH) = (0.5, 0.5) λ. (dg,V, dg,H) = (2.0, 4.0) λ as optional] * Other assumptions are not precluded.   Companies to explain TXRU weights mapping.  Companies to explain beam selection.  Companies to explain number of BS beams |
| **BS Antenna radiation pattern** | TR 38.802 Table A.2.1-6, Table A.2.1-7 |
| **UE Antenna Configuration** | [Panel structure: (M,N,P) = (1,4,2)]   * 2 panels (left, right) with (Mg, Ng) = (1, 2) as baseline * Other assumptions are not precluded   Companies to explain TXRU weights mapping.  Companies to explain beam and panel selection.  Companies to explain number of UE beams |
| **UE Antenna radiation pattern** | TR 38.802 Table A.2.1-8, Table A.2.1-10 |
| **Beam correspondence** | Companies to explain beam correspondence assumptions (in accordance to the two types agreed in RAN4) |
| **Link adaptation** | Based on CSI-RS |
| **Traffic Model** | FFS:   * Option 1: Full buffer * Option 2: FTP model   Other options are not precluded |
| **Inter-panel calibration for UE** | Ideal, non-ideal following 38.802 (optional) – Explain any errors |
| **Control and RS overhead** | Companies report details of the assumptions |
| **Control channel decoding** | Ideal or Non-ideal (Companies explain how it is modelled) |
| **UE receiver type** | MMSE-IRC as the baseline, other advanced receiver is not precluded |
| **BF scheme** | Companies explain what scheme is used |
| **Transmission scheme** | Multi-antenna port transmission schemes  Note: Companies explain details of the using transmission scheme. |
| **Other simulation assumptions** | Companies to explain serving TRP selection  Companies to explain scheduling algorithm |
| **Other potential impairments** | Not modelled (assumed ideal).  If impairments are included, companies will report the details of the assumed impairments |
| **BS Tx Power** | [40 dBm] |
| **Maximum UE Tx Power** | 23 dBm |
| **BS receiver Noise Figure** | 7 dB |
| **UE receiver Noise Figure** | 10 dB |
| **Inter site distance** | 200m |
| **BS Antenna height** | 25m |
| **UE Antenna height** | 1.5 m |
| **Car penetration Loss** | 38.901, sec 7.4.3.2: μ = 9 dB, σp = 5 dB |

Agreement

* For temporal beam prediction, the following options can be considered as a starting point for UE trajectory model for further study. Companies report further changes or modifications based on the following options for UE trajectory model. Other options are not precluded.
  + Option #2: Linear trajectory model with random direction change.
    - UE moving trajectory: UE will move straightly along the selected direction to the end of a~~n~~ time interval, where the length of the time interval is provided by using an exponential distribution with average interval length, e.g., 5s, with granularity of 100 ms.
      * UE moving direction change: At the end of the time interval, UE will change the moving direction with the angle difference A\_diff from the beginning of the time interval, provided by using a uniform distribution within [-45°, 45°].
      * UE move straightly within the time interval with the fixed speed.
    - FFS on UE orientation
  + Option #3: Linear trajectory model with random and smooth direction change.
    - UE moving trajectory: UE will change the moving direction by multiple steps within a~~n~~ time internal, where the length of the time interval is provided by using an exponential distribution with average interval length, e.g., 5s, with granularity of 100 ms.
      * UE moving direction change: At the end of the time interval, UE will change the moving direction with the angle difference A\_diff from the beginning of the time interval, provided by using a uniform distribution within [-45°, 45°].
      * The time interval is further broken into N sub-intervals, e.g. 100ms per sub-interval, and at the end of each sub-interval, UE change the direction by the angle of A\_diff/N.
      * UE move straightly within the time sub-interval with the fixed speed.
    - FFS on UE orientation
  + Option #4: Random direction straight-line trajectories.
    - Initial UE location, moving direction and speed: UE is randomly dropped in a cell, and an initial moving direction is randomly selected, with a fixed speed.
      * The initial UE location should be randomly drop within the following blue area



where d1 is the minimum distance that UE should be away from the BS.

* + - * + Each sector is a cell and that the cell association is geometry based.
        + During the simulation, inter-cell handover or switching should be disabled.

For training data generation

* + - For each UE moving trajectory: the total length of the UE trajectory can be set as T second if it is in time, of set as D meter if it is in distance.
      * The value of T (or D) can be further discussed
      * The trajectory sampling interval granularity depends on UE speed and it can be further discussed.
    - UE can move straightly along the entire trajectory, or
    - UE can move straightly during the time interval, where the time interval is provided by using an exponential distribution with average interval length
      * UE may change the moving direction at the end of the time interval. UE will change the moving direction with the angle difference A\_diff from the beginning of the time interval, provided by using a uniform distribution within [-45°, 45°]
    - If the UE trajectory hit the cell boundary (the red line), the trajectory should be terminated.
      * If the trajectory length (in time) is less than the length of observation window + prediction window, the trajectory should be discarded.
      * At the current stage, the length of observation window + prediction window is not fixed and the companies can report their values.
    - FFS on UE orientation
* Generalization issue is FFS

Agreement

* For temporal beam prediction, further study the following options as baseline performance
  + Option 1a: Select the best beam for T2 within Set A of beams based on the measurements of all the RS resources or all possible beams from Set A of beams at the time instants within T2
  + Option 2: Select the best beam for T2 within Set A of beams based on the measurements of all the RS resources from Set B of beams at the time instants within T1
    - Companies explain the detail on how to select the best beam for T2 from Set A based on the measurements in T1
  + Where T2 is the time duration for the best beam selection, and T1 is a time duration to obtain the measurements of all the RS resource from Set B of beams.
    - T1 and T2 are aligned with those for AI/ML based methods
  + Whether Set A and Set B are the same or different depend on the sub-use case
  + Other options are not precluded.

Agreement

* For dataset generation and performance evaluation for AI/ML in beam management, take the following assumption for LLS as optional methodology

|  |  |
| --- | --- |
| Parameter | Value |
| Frequency | 30GHz. |
| Subcarrier spacing | 120kHz |
| Data allocation | [8 RBs] as baseline, companies can report larger number of RBs  First 2 OFDM symbols for PDCCH, and following 12 OFDM symbols for data channel |
| PDCCH decoding | Ideal or Non-ideal (Companies explain how is oppler) |
| Channel model | FFS:  LOS channel: CDL-D extension, DS = 100ns  NLOS channel: CDL-A/B/C extension, DS = 100ns  Companies explains details of extension methodology considering spatial consistency  Other channel models are not precluded. |
| BS antenna configurations | * One panel: (M, N, P, Mg, Ng) = (4, 8, 2, 1, 1), (dV, dH) = (0.5, 0.5) λ as baseline * Other assumptions are not precluded.     Companies to explain TXRU weights mapping.  Companies to explain beam selection.  Companies to explain number of BS beams |
| BS antenna element radiation pattern | Same as SLS |
| BS antenna height and antenna array downtile angle | 25m, 110° |
| UE antenna configurations | Panel structure: (M, N, P) = (1, 4, 2),   * 2 panels (left, right) with (Mg, Ng) = (1, 2) as baseline * 1 panel as optional * Other assumptions are not precluded     Companies to explain TXRU weights mapping.  Companies to explain beam and panel selection.  Companies to explain number of UE beams |
| UE antenna element radiation pattern | Same as SLS |
| UE moving speed | Same as SLS |
| Raw data collection format | Depends on sub-use case and companies’ choice. |

**Decision:** As per email decision posted on May 25th,

Agreement

* For UE trajectory model, UE orientation can be independent from UE moving trajectory model. FFS on the details.
  + Other UE orientation model is not precluded.

Agreement

* Companies are encouraged to report the following aspects of AI/ML model in RAN 1 #110. FFS on whether some of aspects need be defined or reported.
  + Description of AI/ML model, e.g, NN architecture type
  + Model inputs/outputs (per sub-use case)
  + Training methodology, e.g.
    - Loss function/optimization function
    - Training/ validity /testing dataset:
      * Dataset size, number of training/ validity /test samples
      * Model validity area: e.g., whether model is trained for single sector or multiple sectors
      * Details on Model monitoring and model update, if applicable
  + Others related aspects are not precluded

Agreement

* To evaluate the performance of AI/ML in beam management, further study the following KPI options:
  + Beam prediction accuracy related KPIs, may include the following options:
    - Average L1-RSRP difference of Top-1 predicted beam
    - Beam prediction accuracy (%) for Top-1 and/or Top-K beams, FFS the definition:
      * Option 1: The beam prediction accuracy (%) is the percentage of “the Top-1 predicted beam is one of the Top-K genie-aided beams”
      * Option 2: The beam prediction accuracy (%) is the percentage of “the Top-1 genie-aided beam is one of the Top-K predicted beams”
    - CDF of L1-RSRP difference for Top-1 predicted beam
    - Beam prediction accuracy (%) with 1dB margin for Top-1 beam
      * The beam prediction accuracy (%) with 1dB margin is the percentage of the Top-1 predicted beam “whose ideal L1-RSRP is within 1dB of the ideal L1-RSRP of the Top-1 genie-aided beam”
    - the definition of L1-RSRP difference of Top-1 predicted beam:
      * the difference between the ideal L1-RSRP of Top-1 predicted beam and the ideal L1-RSRP of the Top-1 genie-aided beam
    - Other beam prediction accuracy related KPIs are not precluded and can be reported by companies.
  + System performance related KPIs, may include the following options:
    - UE throughput: CDF of UE throughput, avg. and 5%ile UE throughput
    - RS overhead reduction at least for spatial-domain beam prediction at least for top-1 beam:
      * 1-N/M,
        + where N is the number of beams (with reference signal (SSB and/or CSI-RS)) required for measurement
        + where (FFS) M is the total number of beams
        + Note: Non-AI/ML approach based on the measurement of these M beams may be used as a baseline
      * FFS on whether to define a proper value for M for evaluation.
    - Other System performance related KPIs are not precluded and can be reported by companies.
  + Other KPIs are not precluded and can be reported by companies, for example:
    - Reporting overhead reduction: (FFS) The number of UCI report and UCI payload size, for temporal /spatial prediction
    - Latency reduction:
      * (FFS) (1 – [Total transmission time of N beams] / [Total transmission time of M beams])
        + where N is the number of beams (with reference signal (SSB and/or CSI-RS)) in the input beam set required for measurement
        + where M is the total number of beams
    - Power consumption reduction: FFS on details

Final summary in R1-2205641.

# Agreement in RAN 1 #110

**Agreement**

**The Following updated based on the agreements in RAN 1 #109-e is adopted**

|  |  |
| --- | --- |
| **Parameters** | **Values** |
| **UE distribution** | * ~~FFS~~ 10 UEs per sector/cell for system performance related KPI (if supported) [e.g,, throughput] for full buffer traffic (if supported) evaluation (model inference). * X UEs per sector/cell for system performance related KPI for FTP traffic (if supported) evaluation (model inference). * Other values are not precluded * Number of UEs per/sector per cell during data collection (training/testing) is reported by companies if relevant * ~~More UEs per sector/cell for data generation is not precluded.~~ |
| **UE Antenna Configuration** | * Antenna setup and port layouts at UE: [1,2,1,4,2,1,1], 2 panels (left, right) * ~~[Panel structure: (M,N,P) = (1,4,2)]~~   + ~~panels (left, right) with (Mg, Ng) = (1, 2) as baseline~~ * Other assumptions are not precluded     Companies to explain TXRU weights mapping.  Companies to explain beam and panel selection.  Companies to explain number of UE beams |

**Agreement**

**The Following updated based on the agreements in RAN 1 #109-e is adopted**

|  |  |
| --- | --- |
| **Parameters** | **Values** |
| **UE Speed** | * For spatial domain beam prediction, 3km/h * For time domain beam prediction: 3km/h(optional), 30km/h (baseline), 60km/h (optional), 90km/h (optional), 120km/h (optional) * Other values are not precluded |
| **UE distribution** | * For spatial domain beam prediction:   + Option 1: 80% indoor ,20% outdoor as in TR 38.901   + Option 2: 100% outdoor * For time domain prediction: 100% outdoor |

**Agreement**

* **If UE orientation is modeled, it can be independently modeled from UE moving trajectory model.** 
  + **This is not precluded that UE orientation coupled with UE moving trajectory model.**

**Agreement**

* **Study the following options on the selection of Set B of beams (pairs)** 
  + **Option 1: Set B is fixed across training and inference**
    - **FFS on the beams of Set B**
  + **Option 2: Set B is variable (e.g., different beams (pairs) patterns in each report/measurement during training and/or inference)** 
    - **FFS on fixed or variable number of beams (pairs)**
    - **FFS on the details**
  + **Other options are not precluded.**
  + **FFS on the number of beams (pairs) in Set B**
  + **Note: This does not preclude the alternative that Set B is different from Set A.**

**Agreement**

* **To evaluate the performance of AI/ML in beam management at least for NW side beam prediction, UCI report overhead can be further studied as one of KPI options.** 
  + **FFS: number of UCI reports and UCI payload size**

# Agreement in GTW on 10th Oct

**Working Assumption**

**The following cases are considered for verifying the generalization performance of an AI/ML model over various scenarios/configurations as a starting point:**

* **Case 1: The AI/ML model is trained based on training dataset from one Scenario#A/Configuration#A, and then the AI/ML model performs inference/test on a dataset from the same Scenario#A/Configuration#A**
* **Case 2: The AI/ML model is trained based on training dataset from one Scenario#A/Configuration#A, and then the AI/ML model performs inference/test on a different dataset than Scenario#A/Configuration#A, e.g., Scenario#B/Configuration#B, Scenario#A/Configuration#B**
* **Case 3: The AI/ML model is trained based on training dataset constructed by mixing datasets from multiple scenarios/configurations including Scenario#A/Configuration#A and a different dataset than Scenario#A/Configuration#A, e.g., Scenario#B/Configuration#B, Scenario#A/Configuration#B, and then the AI/ML model performs inference/test on a dataset from a single Scenario/Configuration from the multiple scenarios/configurations, e.g., Scenario#A/Configuration#A, Scenario#B/Configuration#B, Scenario#A/Configuration#B.**
  + - * **Note: Companies to report the ratio for dataset mixing**
      * **Note: number of the multiple scenarios/configurations can be larger than two**
* **FFS the detailed set of scenarios/configurations**
* **FFS other cases for generalization verification, e.g.,**
  + - * **Case 2A: The AI/ML model is trained based on training dataset from one Scenario#A/Configuration#A, and then the AI/ML model is updated based on a fine-tuning dataset different than Scenario#A/Configuration#A, e.g., Scenario#B/Configuration#B, Scenario#A/Configuration#B. After that, the AI/ML model is tested on a different dataset than Scenario#A/Configuration#A, e.g., subject to Scenario#B/Configuration#B, Scenario#A/Configuration#B.**

**Conclusion**

* **For system performance related KPI (if supported) evaluation (model inference), companies report either of the following traffic model:**
  + **Option 1: Full buffer**
  + **Option 2: FTP model with detail assumptions (e.g., FTP model 1, FTP model 3)**

**Agreement**

* **BS antenna configuration:** 
  + **antenna setup and port layouts at gNB: (4, 8, 2, 1, 1, 1, 1), (dV, dH) = (0.5, 0.5) λ**
  + **Other assumptions are not precluded**
* **BS Tx power for evaluation:** 
  + **40dBm (baseline)**
  + **Other values (e.g. 34 dBm) are not precluded and can be reported by companies**
* **UE antenna configuration (Clarification of agreement in RAN 1 #110):** 
  + **antenna setup and port layouts at UE: (1, 4, 2, 1, 2, 1, 1), 2 panels (left, right)**
  + **Other assumptions are not precluded**

**Agreement**

* **For the evaluation of both BM-Case1 and BM-Case2, 32 or 64 downlink Tx beams (maximum number of available beams) at NW side.** 
  + **Other values, e.g., 256, etc, are not precluded and can be reported by companies.**
* **For the evaluation of both BM-Case1 and BM-Case2, 4 or 8 downlink Rx beams (maximum number of available beams) per UE panel at UE side.** 
  + **Other values, e.g., 16, etc, are not precluded and can be reported by companies.**