3GPP TSG RAN WG1 #110bis-e R1-220xxxx

e-Meeting, October 10th – 19th, 2022

Source: Moderator (OPPO)

Title: Summary#2 for other aspects on AI/ML for beam management

Agenda Item: 9.2.3.2

Document for: Discussion and Decision

# Introduction

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

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

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

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

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

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

# Training and deployment of AI/ML model

## Training/inference at UE/NW side

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

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| RAN1#109-e  Agreement  For the sub use case BM-Case1, consider both Alt.1 and Alt.2 for further study:   * Alt.1: AI/ML inference at NW side * Alt.2: AI/ML inference at UE side   Agreement  For the sub use case BM-Case2, consider both Alt.1 and Alt.2 for further study:   * Alt.1: AI/ML inference at NW side * Alt.2: AI/ML inference at UE side   RAN1#110  Agreement  At least for the sub use case BM-Case1 and BM-Case2, support both Alt.1 and Alt.2 for the study of AI/ML model training:   * Alt.1: AI/ML model training at NW side; * Alt.2: AI/ML model training at UE side.   Note: Whether it is online or offline training is a separate discussion.  Working Assumption  Include the following into a working list of terminologies to be used for RAN1 AI/ML air interface SI discussion.   |  |  | | --- | --- | | Terminology | Description | | AI/ML model delivery | A generic term referring to delivery of an AI/ML model from one entity to another entity in any manner.  Note: An entity could mean a network node/function (e.g., gNB, LMF, etc.), UE, proprietary server, etc. | |

In RAN1#110 meeting, there were intensive discussions on whether the AI/model training and inference are at the same node or different nodes (e.g., the AI/ML model can be trained by one node and be used for inference by a node of the opposite side), and the alternatives under discussion were as below:

* Alt.1. AI/ML model training and inference at NW side
* Alt.2. AI/ML model training and inference at UE side
* Alt.3. AI/ML model training at NW side, AI/ML model inference at UE side
* Alt.4. AI/ML model training at UE side, AI/ML model inference at gNB side

After several rounds of discussions and revisions, the following proposal was provided, but not agreed

|  |
| --- |
| (RAN1#110) Proposal 2.1.1-2d: For the sub use case BM-Case1 and BM-Case2, at least support Alt.1 and Alt.2 for AI/ML model training and inference for further study:   * Alt.1. AI/ML model training and inference at NW side * Alt.2. AI/ML model training and inference at UE side * FFS: Alt.3. AI/ML model training at NW side, AI/ML model inference at UE side |

In this meeting, some contributions continue to discuss this issue. The related proposals/observations from the contributions are copied as below:

|  |  |
| --- | --- |
| Huawei[2] | *Proposal 8: Training and inference at the same side is preferred and should be the baseline.* |
| ZTE [3] | *Observation 1: For AI/ML model training and inference, Alt.3 outperforms Alt.1 and Alt.2 in terms of reporting overhead and/or beam prediction performance.*  *Proposal 1: If model transfer is supported in agenda 9.2.1, both Alt.1, Alt.2, and Alt.3 can be supported for further study with potential applicability in different scenarios.* |
| Spreadtrum[4] | *Observation 1: Considering the limitation of UE operation ability and the diversity of antenna array structure, training at UE side may be difficult to complete.* |
| Vivo[5] | *Observation 1: Report overhead may increase dramatically but with less specification impacts for Alt. 1 with enhanced beam pair prediction solution and DL Tx beam prediction solution.*  *Observation 2: Report overhead can be reduced to top-k L1-RSRP and its related Rx beam information, but extra signaling indicated by gNB is needed with limited beam prediction solution for Alt.2 if considering generalization performance.*  *Observation 3: Report overhead is limited for Alt.3 with energy saving solution, but model transfer is needed.*  *Observation 4: The memory storage requirement in NW side seems unaccepted for Alt.4.*  *Proposal 6: For the sub use case BM-Case1 and BM-Case2, at least support Alt.1, Alt.2 and Alt.3 for AI/ML model training and inference:*   * *Alt.1. AI/ML model training and inference at NW side* * *Alt.2. AI/ML model training and inference at UE side* * *Alt.3. AI/ML model training at NW side, AI/ML model inference at UE side* * *Further discuss Alt.4*    + *Alt.4. AI/ML model training at UE side, AI/ML model inference at NW side* |
| IDC[6] | *Observation 1: AI/ML inference/training at NW side (Alt.1) could be a good implementation option as UE implementation is generally limited due to computational power and battery consumption than gNB implementation. However, AI/ML inference/training generally requires more detailed explicit information which leads significant reporting overhead.*  *Observation 2: AI/ML inference/training at UE side (Alt.2) can be limited due to limited computational power and battery consumption at UE implementation, however, UE can easily utilize more information that the UE acquired by measuring SSB/CSI-RS without consuming any reporting overhead.*  *Proposal 1: Support both AI/ML inference/training at NW side (Alt.1) and UE side (Alt.2) for both BM-Case1 and BM-Case2.* |
| OPPO[7] | *Observation 1: For BM-Case1, deploying AI/ML inference at UE side can avoid beam reporting on Set B, therefore resulting in minimum standard impact.*  *Observation 2: For BM-Case2, deploying AI/ML inference at UE side seems more reasonable, otherwise (inference at NW side) there could be overwhelming beam reporting on Set B.*  *Proposal 1: For BM-Case1 and BM-Case2, at least support AI/ML model training and inference at either NW side or UE side.* |
| CATT[11] | *Proposal 1: For the sub use case BM-Case1 and BM-Case2, support the following alternatives for further study:*   * *Alt.1: AI/ML training and inference at NW side;* * *Alt.2: AI/ML training and inference at UE side;* * *Alt.3: AI/ML training at NW side and inference at UE side.* |
| Rakuten[25] | *Proposal 2: Single sided AI/ML (at the gNB side or the UE side) should be considered as baseline.* |
| NVIDIA[26] | *Proposal 3: For the sub use case BM-Case1 and BM-Case2, support both Alt.1 and Alt.2 for the study of AI/ML model training and inference:*   * *Alt.1: AI/ML model training and inference at network side* * *Alt.2: AI/ML model training and inference at UE side* |
| Panasonic[30] | *Proposal 1: Prioritize Alt 1 (AI/ML model training and inference at NW side) and Alt 2 (AI/ML model training and inference at UE side) for further study in this SI.*  *Proposal 2: For DL Tx beam predication, support to study both NW-side and UE-side models* |
| KT[31] | *Observation 1. It seems difficult to apply Alt.1 (i.e., AI/ML model training and inference at NW side) to the initial access procedure due to the resource allocation issue for the measurement reporting.*  *Observation 2. Training AI/ML models on the gNB facilitates more efficient and timely model updating.*  *Observation 3. In terms of signaling overhead and procedures, AI/ML model inference is efficient to operate on the UE side.*  *Proposal 2. For the sub use case BM-Case1 and BM-Case2, at least support the following alternatives for AI/ML model training and inference:*   * *Alt.1. AI/ML model training and inference at NW side* * *Alt.2. AI/ML model training and inference at UE side* * *Alt.3. AI/ML model training at NW side, AI/ML model inference at UE side* |

The view of each company on the above-mentioned alternatives is collected in the following table.

**Mod’s notes:**

* **The following information is based on the contributions submitted to this meeting and the inputs of companies in the last meeting.**
* **If there is no explicit statement/proposal to change the position/preference in its submitted contribution, the input/view of a company on the same issue in FL summaries of the last meeting is assumed to be unchanged. For this case, only company name will be listed without the reference to its tdoc.**
* **If some preference/position/view is not correctly captured, please feel free to correct it**
* **These notes apply to other topics as well.**

|  |  |  |
| --- | --- | --- |
|  | Supported or prioritized | Not supported or down-prioritized |
| Alt.1. AI/ML model training and inference at NW side | Huawei[2], ZTE[3], vivo[5], IDC[6], OPPO[7], CATT[11], NVIDIA[26], Panasonic[30], KT[31], LGE, MTK, NEC, Spreadtrum, DCM, Ericsson, Intel,QC,Apple, SS, Futurewei, Fujitsu, Lenovo, CIACT, Google, Xiaomi, Charter, (26) |  |
| Alt.2. AI/ML model training and inference at UE side | Huawei[2], ZTE[3], vivo[5], IDC[6], OPPO[7], CATT[11], NVIDIA[26], Panasonic[30], KT[31], LGE,MTK, NEC, DCM, Ericsson, Intel, QC, Apple, SS, Futurewei, Fujitsu, Lenovo, CIACT, Google, Xiaomi, Charter, (25) | Spreadtrum[4], |
| Alt.3. AI/ML model training at NW side, AI/ML model inference at UE side | ZTE[3] (if collaboration z is supported) ,vivo[5], CATT[11], KT[31], Apple, Spreadtrum, Ericsson, (7) | LGE, MTK, NEC, NVIDIA, FutureweI, Panasonic, OPPO, QC, FUJITSU, HW, Intel, Samsung, Lenovo, xiaomi, IDC, Charter (16) |
| Alt.4. AI/ML model training at UE side, AI/ML model inference at gNB side | Vivo[5], Apple, (2) | Spreadtrum[4], LGE, CATT, ZTE, MTK, NEC, Lenovo, CAICT, NVIDIA, FutureweI, Panasonic, OPPO, QC, FUJITSU, HW, Ericsson, Intel, Samsung, CMCC, Lenovo, xiaomi, IDC, Charter, (23) |

Based on the above information, we can observe the following:

* Alt.1 is supported by 26 companies
* Alt.2 is supported by 25 companies, not supported by 1 company
* Alt.3 is supported by 7 companies, not supported by 16 companies
* Alt.4 is supported by 2 companies, not supported by 23 companies.

Considering the current status, moderator suggests to deprioritize Alt.4 as the first step since it is only supported by 2 companies but opposed by 23 companies. Meanwhile, some companies commented that Alt.3 depends on whether the mode transfer is supported or not which is being discussed in Agenda item 9.2.1. Thus, the following proposal is given for discussion

###### Proposal 2.1

***Proposal 2.1: For the sub use case BM-Case1 and BM-Case2, at least support Alt.1 and Alt.2 for AI/ML model training and inference for further study:***

* ***Alt.1. AI/ML model training and inference at NW side***
* ***Alt.2. AI/ML model training and inference at UE side***
* ***Regarding whether to support Alt.3 for BM-Case1 and BM-Case2, wait for the conclusion/agreement of Agenda item 2.9.1 on whether to support mode transfer or not***
  + ***Alt.3. AI/ML model training at NW side, AI/ML model inference at UE side***

|  |  |
| --- | --- |
| Company | Comments |
| Google | Support. There could be a typo. 2.9.1 🡪 9.2.1. |
| Nokia | Support. Third bullet is having a typo on “2.9.1”. |
| NEC | Support |
| Spreadtrum | Support |
| LGE | Support |
| Ericsson | Support |
| CATT | Support |
| NTT DOCOMO | Support with fixing the typo. |
| vivo | We are okay for the first two bullets. For the third bullet, our understanding is we have defined model transfer and level z in 9.2.1, and thus it should be no issue to study Alt 3 in this BM agenda. Similarly, we have defined multiple training types considering model transfer and level z in CSI agenda. Hence we are not sure why we cannot do the same here.  Mod: It is up to the group. As agenda item 9.2.1 is discussing the mode transfer and is expected to make decision on the support/prioritization of model transfer. |
| HW/HiSi | Support the first two bullets.  Do not support the third bullet for the general case that would include one-sided and two-sided model. This is because for the one-sided model, there is no need to introduce this commercially complicated case. A general concern from us in this matter is since the gNB and the UE can be from different vendors, the AI model would have to be disclosed to the outside. Additionally, for the UE, it would receive an unseen model from the gNB and it is difficult to control the power consumption, latency and performance.  In our view, the Alt3 can be handled in the scope of the two-sided model, which still is under discussion in 9.2.1.  We would suggest to update the proposal as follows for the one-sided model:  ***Proposal 2.1: For the sub use case BM-Case1 and BM-Case2, for the one-sided AI model, ~~at least~~ support Alt.1 and Alt.2 for AI/ML model training and inference for further study:***   * ***Alt.1. AI/ML model training and inference at NW side*** * ***Alt.2. AI/ML model training and inference at UE side*** * ***~~Regarding whether to support Alt.3 for BM-Case1 and BM-Case2, wait for the conclusion/agreement of Agenda item 2.9.1 on whether to support mode transfer or not~~***   + ***~~Alt.3. AI/ML model training at NW side, AI/ML model inference at UE side~~***   Mod: In my understanding, since only one-sided model was agreed for further study so far, the follow-up discussions are only for one-sided model unless the group agree to support the study of two-sided model. |
| Sony | Support |
| Samsung | Support the proposal. |
| Xiaomi | Support |
| Fujitsu | Support |
| Lenovo | Support |
| MediaTek | Support |
| NVIDIA | Support |
| HW/HiSi | @ moderator: “Mod: In my understanding, since only one-sided model was agreed for further study so far, the follow-up discussions are only for one-sided model unless the group agree to support the study of two-sided model.”  Thanks a lot for the clarification. For the one-sided model, we do not see the need for Alt3 for the reasons we gave in the previous comments. We would then suggest to remove the third bullet.  ***Updated Proposal 2.1: For the sub use case BM-Case1 and BM-Case2, at least support Alt.1 and Alt.2 for AI/ML model training and inference for further study:***   * ***Alt.1. AI/ML model training and inference at NW side*** * ***Alt.2. AI/ML model training and inference at UE side*** * ***~~Regarding whether to support Alt.3 for BM-Case1 and BM-Case2, wait for the conclusion/agreement of Agenda item 2.9.1 on whether to support mode transfer or not~~***   + ***~~Alt.3. AI/ML model training at NW side, AI/ML model inference at UE side~~***   Could proponents of Alt 3 please clarify how to solve the issue of model disclosure and how the UE can guarantee performance, latency and power consumption if the model is trained at the gNB side? |
| Qualcomm | We agree with HW/HiSi and suggest removing the third bullet (Alt. 3). We believe the inclusion of Alt. 3 in the agreement would be premature at this point and the discussion on (Alt. 3) and even the entire proposal can be postponed until feasibility is clarified in 9.2.1. Again, we agree with HW that there are serious underlying implications for Alt. 3 that need to be concretely spelled out before identifying it as an option. This is an active topic being discussed in 9.2.1, and we believe the discussions around this proposal can wait until there’s a crisp understanding of the exact implications. |
| Apple | While alt. 1 can work (e.g., with enhancement with beam reporting), we are not sure how and whether Alt. 2 works given information needed in simulation evaluation may not be available in real deployment, 1) If cell-specific AI model is used: as CSI-RS is configured on a UE specific way, how the UE /UE server is going to collect enough to train the network? Will the UE be pre-loaded with AI models for the cells it is going to connect today or tomorrow? 2) if generic model is used, have we verified generalization performance for AI model, so the UE needs to store only one model or a small number of AI models? I guess the answer is NO. Recall to Alt. 2 work, infra vendors may need to share proprietary information so the UE can train its network properly. There are simply too many unanswered questions with Alt. 2.  As for Alt. 3, a key consideration is generalization and whether potentially very useful information such as antenna configuration, Tx beam angle, etc. can be actually communicated by network to UE. Just as for UE vendors/UE chipset vendors, RX beam design bear proprietary information, the same goes for infra vendors, we wonder what the chance is infra vendors are willing to share such information. Then we are in a dilemma: infra vendors are sensitive to sharing Tx beam info, UE vendors may be reluctant to share Rx beam info. Then some mutually acceptable way needs to be worked out.  If the AI model for UE side inference is trained by network (e.g., in a cell specific way), the infra vendor/operator does not need to reveal anything about Tx beam orientation to anyone, at the UE side the UE does not need to reveal anything to anyone. Then both sides should be more comfortable in the arrangement. We recommend other companies to take a closer look at Alt. 3.  Then as a group, we should not be over-confident with Alt. 2 but at the same time over cautious with Alt. 3. So we suggest agreeing Alt. 1, it takes more to discuss Alt. 2 and Alt. 3. |
| CAICT | Support |
| ZTE | We suggest to postpone the discussion of Alt.3 until a clearer consensus/understanding of model transfer is reached in agenda 9.2.1. Per our understanding, despite the potential risk of proprietary AI/ML model disclosure, Alt.3 may still outperform Alt.2 in terms of model training facilitation and beam prediction performance. Compared with Alt.2, the NW-side model training in Alt.3 could benefit from the powerful computing power and sufficient storage resources of gNB, resulting in well-trained models with better generalization and environmental adaptability. Besides, since gNB has a full understanding of the deployed AI models, some assistance information for AI model input and CSI-RS configuration can be avoided and model LCM are more controllable. |

## Online/offline training

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

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

The related proposals/observations from the contributions are copied as below:

|  |  |
| --- | --- |
| FUTUREWEI[1] | *Observation 1: Given the dynamic nature in the propagation environment, online training (e.g., reinforcement learning) may be a good alternative for AI/ML-based beam management in addition to offline learning approach like supervised learning. As the definition of online training has been approved, it is time to revisit online training as one of the two alternatives of training.*  *Proposal 1: For companies that propose to adopt online training, study the standards impacts, as well as the associated signalling overhead with online training for AI/ML-based beam management.* |
| Huawei[2] | *Observation 6: For NW-side operation mode, model training under online/offline manner is up to implementation.*  *Proposal 9: If an online/offline discussion shall be conducted for the UE-side operation, this discussion should be kept separated from the issue whether data set collection is via air-interface or non-air-interface.* |
| Spreadtrum[4] | *Observation 2: Regarding AI/ML training for BM-Case1 and BM-Case2, offline training should be enough.* |
| OPPO[7] | *Proposal 2: For AI/ML beam management, focus on offline model training at least at current stage.* |
| Fujitsu[12] | *Proposal 1: For the sub use case BM-Case1 and BM-Case2, the study on type of AI/ML model training is suggested to consider the model deployment.*  *Proposal 2: For the sub use case BM-Case1 and BM-Case2, the following type of AI/ML model training is suggested to study:*   * *Online and offline training for NW-side model* * *Offline training for UE-side model* |
| NEC[16] | *Proposal 2: For the trained AI/ML model in offline, study the mechanism of model update (e.g., fine-tuning) based on the online data.*  *Proposal 3: Study the mechanism of online data processing.* |
| NVIDIA[26] | *Observation 1: Offline training may be more feasible for the near future. But in the long run, it is vital that the AI/ML models can learn continuously to adapt to varying environments, site-specific conditions, and heterogenous configurations.*  *Proposal 2: For the sub use case BM-Case1 and BM-Case2, support both Alt.1 and Alt.2 for the study of AI/ML model training:*   * *Alt.1: offline training* * *Alt.2: online training* |
| QC[29] | *Proposal 1: For training of UE-side AI/ML model for beam prediction (BM-Case1 and BM-Case2), focus should be on offline training scenario, in which the development and training of the AI/ML model happens offline without the need to involve 3gpp signaling.* |

Based on the contributions submitted to this meeting and the inputs of the last meeting offline training can be supported by all companies. The controversial part is whether to support online training (i.e., reinforcement learning) or not.

|  |  |  |
| --- | --- | --- |
|  | Supported or prioritized | Not supported or down-prioritized |
| Offline training | All companies |  |
| Online training | Futurewei[1], Fujitsu[12], NVIDIA[26], Sony, Nokia, Google, | Spreadtrum[4], OPPO[7], QC[29], MTK, Ericsson, |

As the definition of online and offline training were not defined, in the last meeting many companies (e.g., LGE, CATT, NEC, NVIDIA, Xiaomi, Panasonic, vivo,) suggest to postpone the discussion until the definition and scope of online training is clearer. In RAN1#110 meeting, a working assumption were made for their definitions. Thus, this agenda item can continue to discuss this issue.

Offline training is supported by all companies and it should be supported for AI-based beam management. The controversial part is whether to support online training or not. Considering there are similar discussion on offline/online training in Agenda item 9.2.1, it is better to avoid the duplicated discussion and the potential confliction. Thus, the following conclusion is proposed for discussion**.**

###### Conclusion 2.2

***Conclusion 2.2: For the sub use case BM-Case1 and BM-Case2, Agenda item 9.2.3.2 focuses on spec impact of the AI-based solution with the assumption of offline training***

* ***Discussion on spec impact of the AI-based solution with the assumption of online training is postponed to wait for the conclusion/agreement of Agenda item 9.2.1 whether online training is supported or not***

|  |  |
| --- | --- |
| Company | Comments |
| Google | Support |
| Nokia | Based on the latest assumptions on online and offline training, the training stage or stages may be transparent to air-interface. Therefore, having any restrictions on possible online training studies may not enable RAN1 to study full potential of AI/ML approaches for BM-case1 and BM-case2. What is important is to list down spec impacts for offline training to see if they are clear to all companies.   |  |  | | --- | --- | | Online training | An AI/ML training process where the model being used for inference) is (typically continuously) trained in (near) real-time with the arrival of new training samples.  Note: the notion of (near) real-time vs. non real-time is context-dependent and is relative to the inference time-scale.  Note: This definition only serves as a guidance. There may be cases that may not exactly conform to this definition but could still be categorized as online training by commonly accepted conventions.  Note: Fine-tuning/re-training may be done via online or offline training. (This note could be removed when we define the term fine-tuning.) | | Offline training | An AI/ML training process where the model is trained based on collected dataset, and where the trained model is later used or delivered for inference.  Note: This definition only serves as a guidance. There may be cases that may not exactly conform to this definition but could still be categorized as offline training by commonly accepted conventions. |   Mod: As confirmed by Taesang, in agenda 9.2.1, there is some discussion/proposal on the prioritization of online and offline training. We should keep consistency with the output of agenda 9.2.1 |
| spreadtrum | Support |
| LGE | We are ok with the proposal |
| Ericsson | Support |
| CATT | We are afraid that whether online training is supported or not can be decided in 9.2.1, since this can be rely on different use cases. But offline training can be as a starting point for our point of view.  Mod: As confirmed by Taesang, in agenda 9.2.1, there is some discussion/proposal on the prioritization of online and offline training. We should keep consistency with the output of agenda 9.2.1 |
| NTT DOCOMO | Fine with the proposal. |
| vivo | OK |
| HW/HiSi2 | ~~Support~~ |
| Sony | Support |
| Xiaomi | Support |
| Fujitsu | Support |
| Lenovo | Support |
| Mediatek | Support |
| NVIDIA | Fine |
| HW/HiSi2 | After more considerations, I updated the previous response and it would firstly be great to have clarified that spec impact for training, regardless if it is on-line or off-line, is only expected for the UE side model. For the gNB side, training is up to implementation. It would be great to hear more views on this.  Mod: By reading the contributions, some companies think there will be some spec impact (e.g. new/enhanced reporting, assistance information). My suggestion is to discuss the impact step by step. For example, at the first step, we can have some progress on high-level and then we can discuss the impact of training at UE side and gNB side, respectively. The similar approach is used for the discussion of inference and model monitoring. |
| Qualcomm | Support |
| Samsung | Ok with the conclusion. |
| NEC | Support |
| CAICT | Support |
| ZTE | Support. In the current stage, we prefer to assume the deployed model has been well trained in offline and can be used directly without any additional online processing or 3gpp signaling. |

# Sub use cases of BM-Case1 and BM-Case2

## General views

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

|  |
| --- |
| **RAN1#109-e**  Agreement  For AI/ML-based beam management, support BM-Case1 and BM-Case2 for characterization and baseline performance evaluations   * BM-Case1: Spatial-domain DL beam prediction for Set A of beams based on measurement results of Set B of beams * BM-Case2: Temporal DL beam prediction for Set A of beams based on the historic measurement results of Set B of beams * FFS: details of BM-Case1 and BM-Case2 * FFS: other sub use cases   Note: For BM-Case1 and BM-Case2, Beams in Set A and Set B can be in the same Frequency Range |

The related proposals/observations from the contributions are copied as below:

|  |  |
| --- | --- |
| Nokia[20] | Observation 15: For inter-cell beam measurements and reporting, it is not clear how the beam prediction in the spatial domain is applied.  Proposal 7: For UE side DL Tx beam prediction with inter-cell beam measurements and reporting, RAN1 shall further study the feasibility of applying beam predictions (BM-Case1 and BM-Case2) across different PCIs or within one PCI.  Proposal 8: For UE side DL Tx beam prediction, Ran1 shall further study group-based beam reporting supported for mTRP operation, including whether Set B measurements can be from two TRPs and UE can report beam pairs from Set A. |
| TCL[21] | Proposal 2: The predictive beam switching shall be discussed in sub use cases of inter-cell beam switching and intra-cell beam switching for latency reduction.  Proposal 4: The subsets of beams at the gNB side and UE side, can be constructed with the assistance of an ML model to reduce the beam training overhead.  Proposal 5: The beam failure detection performance can be enhanced by an AI/ML model based on historical beam measurements. |
| ETRI[22] | Proposal 1. When a beam prediction error occurs, it may be necessary to study whether it can be determined as a beam failure.  Proposal 2. For beam failure recovery according to AI/ML function, it is necessary to study the performance evaluation and specification effect according to comparison with the legacy method. |
| Rakuten[25] | 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. |
| DCM[28] | Proposal 9: Study two-stage beam measurements with top-N predicted beams, since it reduces RS measurement overhead and increases the reliability of beam selection compared to top-1 beam prediction. |

**Mod recommendation**: TBD

## Type of beam prediction

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

|  |
| --- |
| **RAN1#110**  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 |

The related proposals/observations from the contributions are copied as below:

|  |  |
| --- | --- |
| Huawei[2] | *Observation 3: For the beam prediction mechanisms for BM-Case1 and BM-Case2, Alt.1 (DL Tx beam prediction) is a natural replacement of the legacy P1/P2 procedure for Tx beam sweeping, and is compatible to the number/pattern of Rx beams.*  *Observation 4: For the beam prediction mechanisms for BM-Case1 and BM-Case2, the performance gain of Alt.2 (DL Rx beam prediction) may be limited due to a relatively small number of wide Rx beams at UE.*  *Observation 5: For the beam prediction mechanisms for BM-Case1 and BM-Case2, Alt.3 (beam pair prediction) needs NW to be aware of the numbers/patterns of Rx beams and is less flexible in case of varying Rx beams.*  *Proposal 6: For the beam prediction mechanisms for BM-Case1 and BM-Case2, consider Alt.1 (DL Tx beam prediction) as a starting point due to its simplicity and flexibility.* |
| Spreadtrum[4] | *Proposal 2: For sub use cases BM-Case1 and BM-Case2, support Alt3 Beam pair prediction as the predicted beams.* |
| Vivo[5] | *Proposal 5: Study the two AI-based beam prediction solutions for both BM-Case1 and BM-Case2, i.e. enhanced beam pair prediction scheme and DL Tx beam prediction scheme, and considering specification impacts with generalization aspects, such as Set B construction, supported number of Tx/Rx beams, various number of antenna configurations, etc.* |
| IDC[6] | *Proposal 16: DL Rx beam prediction (Alt. 2) should be a part of UE implementation.* |
| OPPO[7] | *Proposal 3: For BM-Case1 and BM-Case2, at least support beam pair prediction (Alt.3) as the key feature of representative sub use cases.* |
| Fujitsu[12] | *Proposal 3: For the sub use case BM-Case1 and BM-Case2, study on DL beam pairs prediction should be prioritized.* |
| Intel[13] | *Proposal1: Beam Pair prediction (Alt-3) should be supported, at least for BM-Case 1 since it can provide large latency and measurement gains for joint P2/P3 procedure* |
| CIACT[17] | *Proposal 1: BM-Case1 should be further refined and clarified according to beam management process.*  *Proposal 2: DL Tx beam prediction and beam pair prediction should be provided higher priority than DL Rx beam prediction.* |
| Nokia[20] | *Observation 2: For DL Tx-Rx beam prediction in the NW side, Set B will be unknown to NW as NW most likely cannot directly control the UE Rx beam selection.*  *Observation 2: For DL Tx-Rx beam prediction in the NW side, Set B will be unknown to NW as NW most likely cannot directly control the UE Rx beam selection.*  *Proposal 2: For BM-Case1 with Set A/B considering Tx-Rx pairs, further discussion may be needed on NW side DL Tx-AoA prediction, UE position information as assistant info to the input of ML model.*  *Observation 3: For UE side TX-RX beam pair inference, it is up to UE to schedule its Rx beam operation for receiving the DL Tx beams. UE can select an Rx beam receiving pattern that is beneficial to its Tx-Rx beam pair prediction.*  *Observation 4: To ensure good prediction performance and maintain the system throughput, the necessary measurement space for DL Tx-Rx beam pair prediction |Set B|\_(Tx-Rx) may increase significantly compared to the measurement space |Set B|\_Tx+|Set B|\_Rx from predicting DL Tx beams and DL Rx beams independently.*  *Observation 5: In Tx-Rx beam pair prediction, the best Rx beam for the non-best Tx beams may need extra Rx beam sweeping or need other extra measurements to be determined.*  *Observation 6: It is unclear what is the performance gain (throughput scaled by overhead, latency ) for predicting the beam pair jointly compared to predicting Tx and Rx independently.*  *Observation 7: For the use case of DL Rx beam prediction, UE needs to report its Rx beam capability and the needed Rx beam sweeping number, which may be different from the UE Rx beam capability max Number of Rx Beam.*  *Proposal 3: For BM-Case1, considering beam types of Set A/B, prioritize Alt.1: DL Tx beam prediction for further study.*   * *RAN1 may consider Alt.2: DL Rx beam prediction and Alt.3: Beam pair prediction as an additional scenario if the benefits are identified in 9.2.3.1.*   *Proposal 11: For BM-Case2 construction of Set A/B, prioritize Alt.1: DL Tx beam prediction for further study.*   * *RAN1 may consider Alt.2: DL Rx beam prediction and Alt.3: Beam pair prediction as an additional scenario if the benefits are identified in 9.3.2.1.* |
| Samsung[27] | *Proposal 8: For predicted beams, Alt 1 (DL Tx beam prediction) is preferred.* |
| Panasonic[30] | *Proposal 3: Prioritize Alt.1 DL Tx beam prediction for further study over Alt.2 Rx beam prediction and Alt 3 Tx-Rx beam pair prediction.* |

In the last meeting, most companies prefer not to include Alt.2 as they think there is no spec impact. Meanwhile, some other companies think there may be some spec impact, e.g., signaling from NW to facilitate the AI/ML model inference at UE side.

By going through the contributions submitted to this meeting, many companies suggest to down-select some out of these 3 alternatives. It seems that Alt.1 and Alt.3 have more supporters whereas Alt.2 has limited supporters.

|  |  |  |
| --- | --- | --- |
| Type of the predicted beams for BM-Case1 and BM-Case2 | | |
|  | Prioritize | Down-prioritize |
| Alt.1: DL Tx beam prediction | Huawei[2], vivo[5], Fujitsu[12], CIACT[17], Nokia[20], Samsung[27], Panasonic[30] |  |
| Alt.2: DL Rx beam prediction |  | IDC[6] (a part of UE implementation), |
| Alt.3: Beam pair prediction | vivo[5], OPPO[7], Fujitsu[12],Intel[13], CIACT[17] |  |

Moreover, the group may have better understanding on the advantages and disadvantages of each alternative as well as their spec impact since we had one more meeting for the study. Thus, it is possible to discuss further down-selection in this meeting. Considering the above information, let’s try to check whether Alt.2 is deprioritized can be accepted all companies or not.

###### Proposal 3.2

***Proposal 3.2: For the sub use case BM-Case1 and BM-Case2, focus on Alt.1 and Alt.3 for the predicted beams for further study with potential down-selection.***

* ***Note: Alt.1 and Alt.3 were agreed in RAN1#110 meeting as below*** 
  + ***Alt.1: DL Tx beam prediction***
  + ***Alt.3: Beam pair prediction (a beam pair consists of a DL Tx beam and a corresponding DL Rx beam)***

As suggested by DCM/QC, the down-selection may be different for UE-side model and network side model, a table is added to collect the views

|  |  |  |
| --- | --- | --- |
| Type of the predicted beams for BM-Case1 and BM-Case2 (UE-side model) | | |
| Alt.1: DL Tx beam prediction |  |  |
| Alt.2: DL Rx beam prediction |  |  |
| Alt.3: Beam pair prediction |  |  |
| Type of the predicted beams for BM-Case1 and BM-Case2 (gNB-side model) | | |
| Alt.1: DL Tx beam prediction |  |  |
| Alt.2: DL Rx beam prediction |  |  |
| Alt.3: Beam pair prediction |  |  |

|  |  |
| --- | --- |
| Company | Comments |
| Google | We think we should focus on Alt1. Rx beam related should be transparent without any spec impact. |
| NTT DOCOMO | We think it is better to down-select for UE side beam prediction and NW side beam prediction, separately. For NW side beam prediction, we prefer to down select Alt1.  Mod: A table is added to collect companies’ view |
| vivo | We think it is not needed to discuss such down-selection at this early stage. These Alts should be studied and evaluated in EVM agenda first. |
| Sony | We think the RX beam prediction might have standard impact for example the period of prediction needs to be coordinated with the CSI-RS PS repetition period. Therefore, we should support Alt. 2. |
| Xiaomi | Support |
| Fujitsu | Support. And it’s not necessary to down select in current stage |
| Mediatek | Alt1 could be the starting point (baseline). Alt3 can also be done. |
| NVIDIA | Alt. 1 should be the baseline. |
| HW/HiSI | Support to focus on Alt1 and Alt3, but we agree with Fujitsu that it is not necessary to down-select between them at this stage. |
| Qualcomm | When we consider the specification impact, Alt. 1 and Alt. 3 may have different implications and feasibility considerations based on whether the AI/ML model is UE-side or gNB-side. That is why the potential down-selection is inevitably a function of at which side the AI/ML model is located. Agree with NTT DOCOMO on this front. |
| Apple | We support Alt. 1 |
| Samsung | Support the proposal in general. For the remaining alternatives, Alt-1 is preferred. Legacy mechanism/implementation already provide a solid performance on the determination of Rx beam. Hence, at least as baseline, Alt-1 should be supported due to its simplicity and robustness in terms of model training/inference. |
| LGE | We support Alt1 |
| NEC | We support Alt.1 as starting point. For Alt.3, we prefer to study beam pair prediction at UE sided because it seems that DL Rx beam is only known to the UE. BTW, there is a typo in the table above, that is, “gNB-sided model” appears twice. |
| CAICT | Fine with FL’s proposal. |
| ZTE | Support to study Alt.1 and Alt.3, with Alt.1 being taken as a baseline. In Alt.3, the beam indication and reporting may lead to UE proprietary information disclosure issues for a NW-side model. Besides, according to our preliminary simulation results, the Tx-Rx beam pairs prediction performs worse that the Tx beam prediction even with a same sampling rate on the whole beam space. |

## Construction of Set A and Set B

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

|  |
| --- |
| **RAN1#109-e**  Conclusion  For the sub use case BM-Case1, consider the following alternatives for further study:   * Alt.1: Set B is a subset of Set A   + FFS: the number of beams in Set A and B   + FFS: how to determine Set B out of the beams in Set A (e.g., fixed pattern, random pattern, …) * Alt.2: Set A and Set B are different (e.g. Set A consists of narrow beams and Set B consists of wide beams)   + FFS: the number of beams in Set A and B   + FFS: QCL relation between beams in Set A and beams in Set B   + ~~FFS: construction of Set B (e.g., regular pre-defined codebook, codebook other than regular pre-defined one)~~ * Note1: Set A is for DL beam prediction and Set B is for DL beam measurement. * Note2: The narrow and wide beam terminology is for SI discussion only and have no specification impact * Note3: The codebook constructions of Set A and Set B can be clarified by the companies.   Conclusion  For the sub use case BM-Case2, further study the following alternatives with potential down-selection:   * Alt.1: Set A and Set B are different (e.g. Set A consists of narrow beams and Set B consists of wide beams)   + FFS: QCL relation between beams in Set A and beams in Set B * Alt.2: Set B is a subset of Set A (Set A and Set B are not the same)   + FFS: how to determine Set B out of the beams in Set A (e.g., fixed pattern, random pattern, …) * Alt.3: Set A and Set B are the same * Note1: Predicted beam(s) are selected from Set A and measured beams used as input are selected from Set B. * Note2: It is up to companies to provide other alternative(s) * Note3: The narrow and wide beam terminology is for SI discussion only and have no specification impact   **RAN1#110**  Agreement  For the sub use case BM-Case1, support the following alternatives for further study:   * Alt.1: Set A and Set B are different (Set B is NOT a subset of Set A) * Alt.2: Set B is a subset of Set A * Note1: Set A is for DL beam prediction and Set B is for DL beam measurement. * Note2: The beam patterns of Set A and Set B can be clarified by the companies.   Agreement  For the sub use case BM-Case2, further study the following alternatives:   * Alt.1: Set A and Set B are different (Set B is NOT a subset of Set A) * Alt.2: Set B is a subset of Set A (Set A and Set B are not the same) * Alt.3: Set A and Set B are the same * Note1: The beam pattern of Set A and Set B can be clarified by the companies. |

The related proposals/observations are copied as below:

|  |  |
| --- | --- |
| Huawei[2] | *Observation 1: For the alternatives of the Set A and Set B relationship under BM-Case 2, Alt.3 (Set A and Set B are the same)*   * *Can inflict compatibility issues with non-AI/ML-based UEs* * *Results into a large beam sweeping overhead during the observation phase* * *May cause unnecessary high interference to cells from neighbor Ues.*   *Proposal 4: For the study of the alternatives of the Set A and Set B relationship under BM-Case 2,*   * *Prioritize the study of Alt.1 (Set A and Set B are different) and Alt.2 (Set B is a subset of Set A).* * *Alt.3 (Set A and Set B are the same) can be optionally used for performance comparison in evaluations.* |
| ZTE[3] | *Observation 2: The number of beams for measurement (i.e., set B) and for prediction (i.e., set A) is related to the trade-off between inference performance and RS overhead for beam measurement.*  *Proposal 3: The sub-sampling based method in Alt.2 can serve as a starting point for the study of spatial domain beam prediction.*  *Proposal 4: The association in reference signals between two sets with different beam settings need to be further studied.*  *Observation 5: The specific construction of the beam set for measurement will have a great impact on the beam training overhead, model complexity and inference performance.*  *Observation 6: If the beam set for measurement and the beam set for prediction are different, the spatial domain beam prediction algorithm may be an essential precondition for the temporal beam prediction study.*  *Proposal 7: Regarding the beam set construction, Alt.3 can be used as a benchmark, while Alt.1 and Alt.2 are deferred until the evaluation of the spatial domain beam prediction in BM-Case1 has achieved sufficient progress.* |
| Spreadtrum[4] | *Proposal 1: For Alt.1 of sub use cases BM-Case1,*   * *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.* |
| IDC[6] | *Observation 3: As using same beamwidth for all channels and signals is a general implementation* *within a frequency range, using a subset of Set A as Set B is a reasonable option if Set A and Set B are utilized in a same frequency range.*  *Observation 4: It is difficult to use a subset of Set A considering different beamwidths for beam management between different frequency ranges.*  *Observation 5: Utilization of wide beam information from a low frequency range has great potential as a low frequency range is more reliable and utilization of wide beam requires much less time and frequency resources for beam management.*  *Proposal 2: Support ‘Set B is a subset of Set A’ when Set A and Set B are utilized in a same frequency range for both BM-Case1 and BM-Case2.*  *Proposal 3: Support ‘Set A and Set B are different’ when Set A and Set B are utilized in different frequency ranges for both BM-Case1 and BM-Case2.*  *Mod: It seems that the cases with different FRs belong to BM-Case3.* |
| OPPO[7] | *Proposal 4: For BM-Case1, Set B is a subset of Set A.*  *Proposal 5: For BM-Case2, Set B and Set A are the same.* |
| LGE[9] | *Proposal #3: For the relation between Set A and Set B of BM-Case2, start from Alt3 to see the feasibility and performance gain of pure TD prediction as an independent approach as SD prediction. After studying this, joint SD and TD prediction (i.e. Alt1 and Alt2) can be studied as a next step.* |
| Sony[14] | *Proposal 3: At least for sub use case 1 and sub use case 2, Set A and Set B are different, Set B with wide beams and Set A with narrow beams.* |
| CIACT[17] | *Proposal 3: The relationship between Set A and Set B in different use cases needs to be clarified in combination with the use case description for further down-selection.* |
| Xiaomi[18] | *Proposal 2: For spatial domain beam prediction, consider set B is a subset of set A with high priority.*  *Proposal 7: For temporal beam prediction, consider set B is same as set A with high priority.* |
| Nokia[20] | *Observation 1: The regular “continuous” wide beam design (i.e., adjacent narrow beams associated to the same wide beam) may not be sufficient to implement a narrow beam prediction with good performance.*  *Proposal 1: For BM-Case1, considering construction of Set A/B, prioritize Alt.2: Set B is a subset of Set A for further studies.*  *• RAN1 may consider Alt1: Set A and Set B are different (Set B is NOT a subset of Set A) as an additional scenario if the benefits are identified in 9.3.2.1.*  *Observation 16: When Set B and Set A are the same in BM-Case2, the beam prediction performance should be the optimum.*  *Proposal 9: In BM-Case2, “Set B and Set A are the same” should be the baseline to study the prediction performance.*   * *FFS relation between K and F with different UE speeds, different channel assumptions, and different measurement periods.*   *Observation 17: For Set B is a subset of Set A, if the selection of Set B is fixed or random over the entire K, then the actual best beam in Set A may or may not be known during K, and the beam prediction performance may degrade.*  *Observation 18: For BM-Case2 Set B is a subset of Set A , for each time instant in K, spatial domain beam prediction or Bayesian optimization can be used to track the best beam over Set A.*  *Observation 19: For BM-Case2 Set B is different to Set A, the actual best beam in Set A may or may not be known during K, and the beam prediction performance may degrade.*  *Observation 20: For BM-Case2 Set B is a subset of Set A , for each time instant in K, spatial domain beam prediction with NN or Gaussian Process can be used the track the best beam over Set A.*  *Proposal 10: In BM-Case2, prioritize studying “Alt.3 Set B and Set A are the same” and “Alt.2 Set B is a subset of Set A”*   * *FFS use cases of Alt.1 Set B and Set A are different.* |
| MTK[23] | *Proposal 1: For BM-Case 1, RAN1 should discuss and define a method or procedure to determine Set B from Set A, when Set B is a subset of Set A and when Set B is different from Set A.*  *Proposal 2: For BM-Case 2, RAN1 should discuss and define method(s) or procedure to determine to determine Set B from Set A, when Set B is a subset of Set A and when Set B is different from Set A.* |

The views are summarized in the following tables:

|  |  |
| --- | --- |
| BM-Case 1 | |
| Alt.1 Set A and Set B are different (Set B is NOT a subset of Set A) | Sony[14], |
| Alt.2 Set B is a subset of Set A | ZTE[3], Spreadtrum[4](for UE-side model), IDC[6] (for the same FR),OPPO[7],xiaomi[18], Nokia[20], |
|  | Spreadtrum[4](up to gNB implementation for NW-side model) |

|  |  |
| --- | --- |
| BM-Case 2 | |
| Alt.1 Set A and Set B are different (Set B is NOT a subset of Set A) | Huawei[2] , IDC[6] (for the same FR), Sony[14], |
| Alt.2 Set B is a subset of Set A (Set A and Set B are not the same) | Huawei[2], IDC[6] (for different FRs), |
| Alt.3 Set A and Set B are the same | ZTE[3], OPPO[7], LGE[9], xiaomi[18],Nokia[20], |

From the above 2 tables, we can see that only a limited number of contributions discussed the further down-selection on the alternatives of BM-Case1 and BM-Case2. Meanwhile, some companies changed their preferences. Let’s try to collect companies’ views on whether some alternative(s) should be prioritized/deprioritized in this stage. If so, which alternative(s) is to be down-selected.

###### Check views 3.3

|  |  |  |
| --- | --- | --- |
| BM-Case1 | | |
|  | Prioritized | Deprioritized |
| Alt.1 Set A and Set B are different (Set B is NOT a subset of Set A) |  |  |
| Alt.2 Set B is a subset of Set A | vivo, NVIDIA, NEC, ZTE |  |
| BM-Case2 | | |
|  | Prioritized | Deprioritized |
| Alt.1 Set A and Set B are different (Set B is NOT a subset of Set A) | HW/HiSi |  |
| Alt.2 Set B is a subset of Set A (Set A and Set B are not the same) | Vivo, HW/HiSi, ZTE |  |
| Alt.3 Set A and Set B are the same | DCM, NVIDIA, ZTE | [HW/HiSi] |

* Whether down-selection is needed in this meeting? (Y/N)
* If yes, which alternative(s) is suggested to be deprioritized or prioritized for Case 1 and Case 2, respectively? (please provide inputs in the above table if any)

|  |  |  |
| --- | --- | --- |
| Company | Y/N | Comments |
| Google | N | We think it is unnecessary for down-selection. Both Alt1 and Alt2 look to be valid use cases. |
| CATT | N | We don’t think it’s necessary for down-selection in this meeting. All Alts can be further studied under BM-Case1 and BM-Case2. |
| Vivo |  | We support using set B is a subset of set A for both Case 1 and Case 2 considering overhead reduction and beam management performance. No down-selection is also fine. Both Alt 1 and Alt 2 can be studied and evaluated first. |
| Sony | N | It is up to implementation for defining the set B. At current stage we should study the standard impact of Case 1 and Case 2 instead of down-selecting different detailed options. |
| Xiaomi |  | We prefer to use Alt 2 for BM case 1 and Alt 3 for BM case 2. |
| MediaTek | N | These cases seem fine. For BM Case 2, Alt 3 could be the baseline to start with. |
| HW/HiSi | N | For BM-Case 2, we think that Alt3 only can be used for performance evaluation and comparison. Alt 3 is in our view not suitable for practical implementation, die to its large overhead during the observation phase. Therefore, we think that from start Alt1 and Alt2 have to be considered also for BM Case 2. |
| Qualcomm | N | No need to down-select at this point. |
| Samsung | N | In our view, there is no need to do the down-selection in A.I. 9.2.3.2. It is more preferable that, these alternatives are down-selected or recommended based on the evaluation results in A.I. 9.3.2.1. |
| LGE |  | No need to down-select, but for BM-Case2, we prefer to have step-by-step approach for more organized discussion/observation, i.e. observe pure TD-prediction(Alt3) first, then observe joint TD-SD prediction (Alt1/2). This does not mean to rule-out specific alt(s). |
| NEC |  | For BM-Case2, we support Alt.3 as starting point. And we are open to Alt.1 and Alt.2. |
| ZTE | N | We prefer not to do any down-selection at the current stage, especially for BM-Case2. Alt.2 and Alt.3 in BM-Case2 should be further evaluated to see if sufficient gains can be achieved over the non-AI methods in terms of RS overhead, reporting overhead, and beam prediction accuracy. |

## Details for Set B

### Refinement of Set B

The related proposals/conclusions are copied as below

|  |  |
| --- | --- |
|  |  |
| Samsung[27] | *Proposal 1: For the sub use case BM-Case1, consider to define Set C for AI/ML inference at NW side.*   * *Set C consists of the beams reported by UE from Set B.*   *Proposal 4: For the sub use case BM-Case2, consider to define Set C for AI/ML inference at NW side.*   * *Set C consists of the beams reported by UE from Set B.* |
| QC[32] submitted to EVM agenda item | *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)*   *To clarify Proposal 2 using an example, if there are beams transmitted by gNB and UE measures gNB beams, and reports gNB beams, the following describe Alt. 1 and Alt. 2:*   * *Alt. 1: For both UE-side and gNB-side AI/ML models, Set B is the set of measured beams.* * *Alt. 2: For UE-side AI/ML model* *Set B is the set of measured beams and for gNB-side AI/ML model Set B is the set of “reported” beams whose measurements are “available” at gNB side.* |

###### Clarification of Set B 3.4.1

Two companies (SS[27], QC[32]) point out that the beams used for AI/ML model inputs may be different from the beams that UE measures and suggest to introduce a new set (e.g., Set C) or make some clarification on set B. By reading the contributions submitted to Agenda item 9.2.3.2 and 9.2.3.1, it seems most companies assume the beams of Set B are the inputs of AI/ML model (i.e., Alt.2 of QC[32]). It would be beneficial to clarify which interpretation/definition of Set B is the common understanding.

Companies are encouraged to provides views on the interpretation/definition of Set B. The alternatives from QC[32] are used as the starting point to collect the views. Based on the inputs/progress, some proposal or conclusion may be suggested later.

|  |  |  |
| --- | --- | --- |
| Set B | Support | Not support |
| Interpretation 1: Set B is a set of beams, whose measurements are performed (for prediction of set A) | Samsung, LGE(for UE-side model), ZTE | Google |
| Interpretation 2: Set B is a set of beam whose measurements are available as inputs of the AI/ML model (for prediction of set A) | Google, DCM, vivo，Fujitsu, NVIDIA, HW/HiSi, LGE(for NW-side model), NEC |  |
| If Interpretation 1 is preferred, whether to introduce a new Set C whose measurements are available as inputs of AI/ML model | Samsung, ZTE |  |

Companies can provide detailed inputs (if any) in the following table.

|  |  |
| --- | --- |
| Company | Comments |
| Xiaomi | We suggest the following update for clear  Interpretation 2: Set B is a set of beam whose measurements are performed or reported ~~available~~ as inputs of the AI/ML model (for prediction of set A) |
| Samsung | This clarification is very important for the case that AI/ML model inference at gNB side. In our view, gNB can provide a set of beams (e.g., set B) for beam measurement. However, the number of beams and/or the corresponding measurement results may be filtered/quantized by UE beam reporting mechanism (e.g., only the results of top-4 beam are reported). Hence, it is better to introduce set C which means the set of beams and the corresponding results that can actually obtained by gNB. |
| LGE | Based on previous agreements, our understanding is that SD prediction is performed on measurement Set B. So for UE-side DL beam prediction, Interpretation 1 seems correct. And for NW-side DL beam prediction, Interpretation 2 seems more correct since DL measurement is performed by UE. In summary,   * For UE-side DL beam prediction: Interpretation 1 * For NW-side DL beam prediction: Interpretation 2 |
| ZTE | Support Samsung's view. We prefer not to constrain that all measured beams should be input to the AI model. For a NW-side model, it is beneficial that UE only report some of the measured beams with larger measured RSRP to reduce reporting overhead. Besides, reporting some beams with very small RSRP may not meaningful for AI inference. Take BM-Case 2 as an example, if set A and set B are the same, it is possible that UE measured all beams in each past time instance and only report one best beam with the largest RSRP to gNB, which is used as the AI model input. In this way, the reporting overhead and model complexity can be greatly reduced. |
|  |  |
|  |  |
|  |  |

### Beam pattern for Set B

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

|  |
| --- |
| **RAN1#110**  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 |

The related proposals/observations are copied as below:

|  |  |
| --- | --- |
| Huawei[2] | *Proposal 2: For the study of AI/ML model input, consider a fixed beam as a starting point.* |
| Vivo[5] | *Proposal 2: Support to study specification impact on Set B selection with semi-random beam subset selection scheme which can provided comparable gain to fixed scheme but with higher flexibility and better generalization performance.* |
| Lenovo[15] | *Proposal 1: Selection of beams for Set B should allow for variable beams, i.e., different beams (pairs) patterns during training and/or inference.* |
| TCL[21] | *Proposal 3: Some patterns can be designed for the input set B of beam prediction in spatial domain.*   * *A fixed pattern;* * *A random pattern.* |

When Set B is a subset of Set A, there are different alternatives on how to determine the beam pattern of Set B and the corresponding views are summarized as below:

|  |  |
| --- | --- |
| Beam pattern for Set B if Set B is a subset of Set A | |
| Option 1: Set B is fixed across training and inference | Huawei[2], Lenovo[15], TCL[21], |
| Option 2: Set B is variable | vivo[5] (Semi-random pattern), Lenovo[15], TCL[21], |

According to the tdocs, some companies suggest to prioritize some beam pattern of Set B. Meanwhile, in EVM session, there are also many tdocs showing the evaluation results for different alternatives. Duplicated discussions in the two sub agenda items should be avoided.

**Moderator recommendation**: In order to avoid the duplicated discussion, further down-selection (if any) on the beam pattern of Set B is discussed in EVM session (Agenda item 9.2.3.1). The spec impact of any given beam pattern of Set B, if any, is discussed in this session (Agenda item 9.2.3.2).

|  |  |
| --- | --- |
| Company | Comments |
| Google | We think any down-selection is better to be handled in this agenda, as this could have spec impact. Agenda item 9.2.3.1 is used to define evaluation methodology and to collect evaluation results. |
| vivo | We think it is better to evaluate and study these options in EVM agenda first. The study would need consider generalization performance. |
| Sony | We suggest to discuss the spec impact in steading going to the details of beam pattern design. |
| HW/HiSi | Ok with the moderator’s suggestion |
| Qualcomm | Agree with moderator’s suggestion. |
| Xiaomi | Agree with moderator’s suggestion. |
| LGE | Support |
| NEC | Agree |
| CAICT | Agree |
| ZTE | Agree |

## Input of BM-Case1 and BM-Case2

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

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

The related proposals/observations are copied as below:

|  |  |
| --- | --- |
| FUTUREWEI[1] | *Observation 2: Assistance information may come with additional cost like signalling overhead and there is usually a trade-off between performance gain and the associated overhead.*  *Proposal 2: When assistance information is used as input, study its performance gain vs. the standards impacts and overhead.* |
| Huawei[2] | *Proposal 1: For the BM-Case 1 study of the AI/ML model input,*   * *Alt.1 (Only L1-RSRP for Set B) should be studied with high priority.* * *Alt.2 (L1-RSRP for Set B and assistance information), if studied, should preclude assistance information that requires the disclosure of propriety information to the opposite node.* * *Alt.4 (L1-RSRP for Set B and DL Tx and/or Rx beam ID) can be studied if benefits are justified by evaluation*   *Proposal 3: For the BM-Case 2 study of the AI/ML model input,*   * *Alt.1 (Only L1-RSRP for Set B) should be studied with high priority.* * *Alt.2 (L1-RSRP for Set B and assistance information), if studied, should preclude assistance information that requires the disclosure of propriety information to the opposite node.* * *Alt.3 (L1-RSRP for Set B and DL Tx and/or Rx beam ID) can be studied after benefits are justified by evaluation.* |
| ZTE[3] | *Observation 3: Much of the assistance information mentioned in Alt.2 is proprietary information of the gNB or UE, which may be difficult to be obtained and shared with another vendor.*  *Proposal 5: Focusing the discussion on Alt.1 and Alt.4 with only L1-RSRP measurement and the corresponding beam ID being taken into account for the AI input would be a good starting point.* |
| Spreadtrum[4] | *Proposal 3: Whether to choose Alt 1 or Alt 4 needs further discussion according to the beam pattern selection.* |
| Vivo[5] | *Proposal 1: Regarding to BM-Case1 and BM-Case 2, at least prioritize following AI input information for further study on specification impact:*   * *L1-RSPR measurement based on Set B* * *Corresponding DL Tx beam pointing angle/ID* * *Corresponding DL Rx beam pointing angle/ID* * *Expected Tx and/or expected Rx beam angle/ID* * *Further discuss other information, such as Tx and/or Rx beam shape information, 3dB beam-width, etc.* |
| IDC[6] | *Observation 6: ‘Only L1-RSRP measurement based on Set B’ is not clear enough as the alternative does not provide any beam related information.*   * *If ‘Only L1-RSRP measurement based on Set B’ means that L1-RSRP measurements are provided in a fixed order, in our view, the input is not ‘Only L1-RSRP measurement based on Set B’.* * *Reporting L1-RSRP measurements in a fixed order is indicating L1-RSRP measurement with implicit beam related information.*   *Proposal 5: Companies supporting the alternative should provide more details for predicting L1-RSRP values without any beam information.*  *Observation 7: ‘L1-RSRP measurement based on Set B and the corresponding DL Tx and/or Rx beam ID’ can be a baseline option as AI/ML model can predict RSRP measurements with Tx and Rx beam IDs which are not provided.*  *Proposal 6: Support ‘L1-RSRP measurement based on Set B and the corresponding DL Tx and/or Rx beam ID’ as a baseline.*  *Proposal 7: Additional information such as TRP IDs and Panels IDs should be considered.*  *Proposal 8: ‘CIR based on Set B’ can be considered as an alternative only for beam management based on FR1 information.* |
| Google[8] | *Proposal 1: For spatial domain beam prediction, support Alt3 (CIR based on set B).*  *Proposal 2: For spatial domain beam prediction, support to add CIR+L1-SINR as one alternative, where the L1-SINR can be used to reflect the interference level for the CIR measurement.*  *Proposal 7: For time-domain beam prediction, support to add CIR measurement based on set B as one alternative.*  *Proposal 8: For time-domain beam prediction, support to add CIR+L1-SINR as one alternative, where the L1-SINR can be used to reflect the interference level for the CIR measurement.* |
| LGE[9] | *Proposal #1: For the UE AI/ML input, Alt2 can be considered. For the assist information for input, output, training, and inference, consider to express Set A and Set B beams on a pre-defined or configured beam grid.* |
| Ericsson[10] | *Observation 1 The feasibility of defining a meaningful TX/RX beam shape information for beam prediction is questionable.*  *Proposal 1 Assistance information related to “beams” should focus on information related to NW antenna/beam configuration ID or UE antenna/beam configuration ID*  *Proposal 2 Prioritize assistance information that can be obtained with low standardization effort, such as UE position information*  *Proposal 3 Study assistance information that captures dynamic UE movement (e.g. using sensors)* |
| CATT[11] | *Proposal 2: For the sub use case BM-Case1 and BM-Case2, study the following alternatives for AI/ML input:*   * *Alt.1: Only L1-RSRP measurement based on Set B;* * *Alt.2: L1-RSRP measurement based on Set B and the corresponding DL Tx and/or Rx beam ID* * *Alt.3: L1-RSRP measurement based on Set B and assistance information.*   + *FFS: Assistance information other than beam ID* |
| Sony[14] | *Proposal 4: At least for sub use case 1, support CIR as the AI/ML model input.* |
| Lenovo[15] | *Proposal 2: Assistance information for AI/ML input should be carefully studied considering the availability of different kinds of assistance information for UE-centric or NW-centric AI/ML inference.* |
| Xiaomi[18] | *But for Alt 3, it needs to define a new measurement quantity which will introduce much more additional workload compared to exist L1-RSRP. Thus we prefer to consider it with low priority (Mod note: for case 1)* |
| Nokia[20] | *Observation 8: For BM-Case 1 model input, the Set B L1-RSRP measurements are needed.*  *Observation 9: For the NW side model with DL TX beam prediction, L1-RSRP measurements of Set B are sufficient for ML model input.*  *Observation 10: For the UE side model with DL Tx beam or Tx-Rx beam pair prediction, if NW-UE collaboration level z is not considered, then Set B L1-RSRP and assistant info will be needed as ML model input for Set A and Set B generalization.*  *Observation 11: For UE side DL Tx beam or DL Tx-Rx beam pair prediction, the DL Tx beam indexes or CRI with certain mapping for Set A/B are needed for the UE.*  *Observation 12: In BM-Case1 model input, Alt. 4 should be merged with Alt.2 as beam ID is covered in both alternatives.*  *Proposal 4: Regarding the sub-use case BM-Case1, further study the following alternatives for AI/ML input:*   * *Alt.2: L1-RSRP measurement based on Set B and assistance information*   + *FFS: Assistance information. The following were mentioned by companions in the discussion:  Tx and/or Rx beam shape information (e.g., Tx and/or Rx beam pattern, Tx and/or Rx beam boresight direction (azimuth and elevation), 3dB beamwidth, etc.), expected Tx and/or Rx beam for the prediction (e.g., expected Tx and/or Rx angle, Tx and/or Rx beam ID for the prediction), UE position information, UE direction information, Tx beam usage information, UE orientation information, gNB panel array parameters (bearing angle, mechanical downtilt, slant angle), etc.*   *Proposal 12: Regarding the sub-use case BM-Case2, further study the following alternatives for AI/ML input:*   * *Alt 2: L1-RSRP measurement based on Set B and assistance information*   + *FFS: Assistance information. The following were mentioned by companies in the discussion:, Tx and/or Rx beam angle, position information, UE direction information, positioning-related measurement (such as Multi-RTT), expected Tx and/or Rx beam/occasion for the prediction (e.g., expected Tx and/or Rx beam angle for the prediction, expected occasions of the prediction), Tx and/or Rx  beam shape information (e.g., Tx and/or Rx beam pattern, Tx and/or Rx beam pointing angles beam boresight directions (azimuth and elevation), 3dB beamwidth, etc.) , increase ratio of L1-RSRP for best N beams, UE orientation information* |
| TCL[21] | *Proposal 1: The UE position information is not necessary for predictive beam switching.* |
| MTK[23] | *Proposal 3: RAN1 will study on the details and advancement of UE’s beam-related L1-RSRP report.*  *Proposal 4: Discussions and agreements are needed to prioritize and down-scope alternatives of UE assistance information.* |
| Apple[24] | *Observation 2: the Tx analog beam information is already embedded in the training data. Whether additional information about Tx beams such as Tx beam shape and Tx beam angle can be useful, or concepts such as Tx beam shape and/or Tx beam orientation can be used in practice need further study.*  *Observation 3: conventionally Rx beam design is transparent to network operation, AI/ML aided/enabled beam management does not need to depart from that. Whether additional information about Rx beams such as Rx beam shape and Rx beam angle can be useful, or concepts such as Rx beam shape and/or Rx beam orientation can be used in practice need further study.* |
| NVIDIA[26] | *Observation 2: Evaluation results show that by using L1-RSRP measurement based on Set B of beams, the AI/ML-based algorithm can achieve performance comparable to that of exhaustive beam search in Set A of beams.*  *Proposal 4: For BM-Case 1, at least support L1-RSRP measurement based on Set B of beams as AI/ML model input.*  *Proposal 5: Comprehensive evaluation results showing convincing performance gains is needed to nail down the essential assistance information needed for the spatial-domain DL beam prediction.*  *Observation 3: Evaluation results show that by using historical optimal index, the AI/ML-based algorithm can satisfactorily yield optimal beam index prediction for future time instances.*  *Proposal 6: For BM-Case 2 (temporal DL beam prediction), at least support using historical optimal beam index based on Set B of beams as AI/ML model input.*  *Proposal 7: Comprehensive evaluation results showing convincing performance gains is needed to nail down the essential assistance information needed for the temporal DL beam prediction.* |
| DCM[28] | *Proposal 2: Encourage companies to bring up their views on what deployment information (e.g., beam information or Rx beam ID) can be exchanged between UE and gNB, before discussing the signalling mechanism of assistance information.* |
| QC[29] | *Observation 1: Based on simulation results in [3], for spatial domain beam prediction (BM-Case1), assistance information from gNB about gNB beam boresight directions and information about gNB antenna array structure is beneficial in boosting spectral efficiency across Ues.* |

###### Clarification 3.5.1

In the contributions submitted to this meeting and discussions in previous meetings, there are different understandings on the alternatives of “Only L1-RSRP measurement based on Set B” (Alt.1 for Case1/2) and “L1-RSRP measurement based on Set B and the corresponding DL Tx and/or Rx beam ID” (Alt.4 for Case1, Alt.3 for Case2). In order to facilitate the discussion, let’s assume the AI/ML model will predict 16 beams based on the measurements of 4 beams in Set B and **a fixed pattern (e.g., beam ID 0, 4, 8, 12)** is used for Set B (e.g., Set B is fixed across training and inference) and K1 L1-RSRP measurement results are ordered according to the information about beam ID as the AI/ML input.

* Case X1: The dimension of AI/ML model inputs are 4, where the 4 L1-RSRP is ordered according to the information about beam ID. For example, the i-th element of the input is corresponding to the beam 4\*I (i=0,4,8,12)
* Case X2: The dimension of AI/ML model inputs are 16, where only 4 elements in the positions corresponding to the measured beams are the measured/reported L1-RSRP and other elements are set as zeros or NaN. For example, the i-th element of the input are measured/reported L1-RSRP (i=0,4,8,12)

Companies have different views on the above cases. Let’s take Case X1 as the example

* Some companies think Case X1 belongs to the alternative “Only L1-RSRP measurement based on Set B” since the beam ID information is not explicitly used as input.
* On the contrast, some other companies think Case X1 belongs to the alternative “L1-RSRP measurement based on Set B and the corresponding DL Tx and/or Rx beam ID” since the order of the L1-RSRP measurement results as the input is determined by the beam ID information implicitly.

In order to facilitate the further discussion and progress, it would be beneficial to have a common understanding within the group. Thus, moderator suggests to clarify the alternatives and make some conclusion (if possible) to avoid the ambiguity.

Companies are encouraged to provides views on Case X1, Case X2. Based on the inputs/progress, some proposal or conclusion may be suggested later.

|  |  |  |
| --- | --- | --- |
| **For Option 1: Set B is fixed across training and inference** | | |
|  | Only L1-RSRP  (Alt.1 for Case 1/2) | L1-RSRP + DL beam ID  (Alt.4 for Case1, Alt.3 for Case2) |
| Case X1 | Google, DCM, Xiaomi, NEC, ZTE | CATT，Fujitsu, Samsung |
| Case X2 |  | CATT, DCM, Fujitsu,Xiaomi, Samsung, NEC, ZTE |

Companies are also invited to share their views on the following issues:

* Whether to further split L1-RSRP + DL beam ID (Alt.4 for Case1, Alt.3 for Case2) into two sub- categorize
  + Cat1: L1-RSRS + implicit DL beam ID
  + Cat2: L1-RSRS + explicit DL beam ID
* If yes, how to define Cat1 and Case2, especially the definition of “implicit” and “explicit”

|  |  |
| --- | --- |
| Company | Comments |
| CATT | We think Case X1 and X2 are both belong to implicit beam ID. For explicit beam ID, the corresponding beam ID value needs to be as additional inputs for AI/ML model. |
| NTT DOCOMO | In addition to case X1 and case X2, the following case X3 also can be considered   * Case X3: The dimension of AI/ML model inputs are 8, where 4 elements correspond to the beam ID and the other 4 elements correspond to L1-RSRP associated with the beam ID.   Mod: If I understand correctly, the 4 elements for beam IDs are useless for the fixed beam case. The reason is that the beam IDs are the same for all training data and test data. Companies are invited to share views on Case X3. |
| Vivo | We are not sure whether it is needed to discuss these cases and the above split in this agenda item. This agenda item is about specification impact. Any input requiring beam ID information (explicitly or implicitly) is same from specification perspective. How to formulate/translate the information into the exact AI model input is implementation.  Mod: Companies have different views. Let’s hear more views. |
| Apple | This is related to the discussion on set B, i.e. whether set B is the set of beams with actual beam measurements or set of beams as inputs to the AI inference model. |
| Xiaomi | We think case X1 belongs to only L1-RSRP since there is no accurate beam ID for each L1-RSRP. But case X2 belongs to L1-RSRP + beam ID. |
| Samsung | Either sub-categorization or clarification work for us. The key point is to better differentiate the alternative of ‘Only L1-RSRP’ and ‘L1-RSRP + DL beam ID’. In terms of the X3 mentioned by DOCOMO, it belongs to Cat2. |
| LGE | Similar view with Vivo. There is no strong motivation to refine AI/ML model input/output in this agenda. |
| NEC | Agree with VIVO, that is, this problem is implementation. |
| CAICT | Same view as Samsung. |
| ZTE | We think the differentiation of implicit and explicit beam ID depends on whether the associated beam ID is input to the AI model. Besides, this is related to the model implementation issue and both would have similar spec impacts. |

###### (Closed) Conclusion 3.5.2

Another controversial part for AI/ML input is regarding the assistance information. The views are quite diverging. Many companies thought the contents of assistance information are not clear and request the proponents to disclose detailed assistance information and how to use it. Meanwhile, some companies mentioned or proposed some types of assistance information, e.g., the examples in the following table. Please note that the table is not completed and just used for example. Moreover, for each example in the above table, there are some opposing companies.

|  |  |
| --- | --- |
| Assistance information | |
| vivo[5] | *Corresponding DL Tx beam pointing angle/ID*  *Corresponding DL Rx beam pointing angle/ID*  *Expected Tx and/or expected Rx beam angle/ID* |
| LGE[9] | *‘relative’ information between Set A and Set B, e.g. Set B beams on a beam grid of potential candidates of Set A.* |
| Ericsson[10] | *NW antenna/beam configuration ID or UE antenna/beam configuration ID*  *UE position/direction/orientation information* |
| QC[29] | *Information about gNB beam boresight directions*  *Information about gNB antenna array structure* |

In order to make a smooth progress on the selection of assistance information (if any), moderator feels it is better to agree some principle so that the discussion/down-selection can be more focused. Many companies (including NW vendors and UE/chipset vendors) raised concern that many types of assistance information are proprietary/privacy and cannot be shared among vendors. Not to disclose proprietary/privacy information is widely accepted as a best practice in 3GPP so far. Meanwhile, the performance and spec impact also should be evaluated. Thus, the following conclusion is suggested for discussion:

***Conclusion 3.5.2:*** ***For the determination/selection of assistance information (if supported),***

* ***The proprietary/privacy information should not be disclosed***
* ***The performance and specification impact should be considered***
  + ***Assistance information can be considered when there is some corresponding evaluation result(s).***

|  |  |
| --- | --- |
| Company | Comments |
| Google | Support |
| Nokia | This is ok |
| NEC | Support |
| Spreadtrum | Support |
| LGE | Principle is fine. The last sub-bullet seems too restrictive, prefer to delete it. |
| Ericsson | Support |
| CATT | Support |
| NTT DOCOMO | Support the proposal. |
| vivo | Support in principle. We want to further stress that generalization capability and performance is also an important aspect. Hence we suggest the following rewording for the second bullet.   * ***The performance, generalization and specification impact should be considered*** |
| HW/HiSi | Fine in principle, but can the intention of the second bullet and its sub-bullet be clarified? Is the intention that spec impact is considered as soon as there are evaluation results available? In our view, the spec impact should only be considered if there is a potential gains,  ***Conclusion 3.5.2:*** ***For the determination/selection of assistance information (if supported),***   * ***The proprietary/privacy information should not be disclosed*** * ***The performance and specification impact should be considered***   + ***Assistance information can be considered when there is some corresponding evaluation result(s) that indicate(s) a performance gain over not using assistance information.*** |
| Samsung | Agree the conclusion in general. The disclose of proprietary/privacy information (e.g., beam angle/width) is not acceptable to us. |
| Xiaomi | We are fine with the conclusion |
| Fujitsu | Support |
| Lenovo | Support |
| Mediatek | Support |
| Qualcomm | We believe the prospect of assistance information being “deemed” proprietary should not dissuade companies from evaluating the efficacy of assistance information in their simulations. Is there a clear definition or understanding of proprietary information? Is there a consensus among companies on what exactly is proprietary and what is not? We believe the sequence of events should be as follows: 1) Companies explore and evaluate the efficacy of assistance information in their simulations and then if performance gains are justified 2) debate about the feasibility of signaling assistance information from the viewpoint of revealing proprietary information, rather than the other way round. As mentioned before, there is a precedent to this in 3gpp. gNB beam shape information was “deemed” proprietary but in Rel-17 positioning indication of beam shape information from gNB to UE is already supported in Spec, due to performance gains observed in evaluations.  So, we suggest rewording the proposal to the following:  ***Conclusion 3.5.2:*** ***For the determination/selection of assistance information (if supported),***   * ***The performance and specification impact should be considered*** * ***The prospect of revealing proprietary/privacy information should be considered.*** |

###### Conclusion 3.5.2b

The version discussed in GTW session is copied for further discussion.

***Conclusion 3.5.2b:*** ***For the determination/selection of assistance information (if supported), e.g., for data collection, AI/ML model inputs and other components of LCM***

* ***The proprietary/privacy information should not be disclosed***
* ***The performance, generalization and specification impact should be considered***
  + ***~~Assistance information can be considered when there is some corresponding evaluation result(s).~~***

|  |  |
| --- | --- |
| Company | Comments |
| nvidia | Support |
| HW/HiSi | We prefer the original proposal without the last sub-bullet.  The yellow part “***e.g., for data collection, AI/ML model inputs and other components of LCM***”, may at this stage cause more confusion than it resolves. This could be part of a later discussion, for example whether different assistance information is intended for different purposes. Could be god to remove the  The “generalization” of the last bullet can be removed in our view. It seems that performance includes generalization behavior already. Could it be clarified, why it is suggested to explicitly add “generalization”?  ***Conclusion 3.5.2b:*** ***For the determination/selection of assistance information (if supported), ~~e.g., for data collection, AI/ML model inputs and other components of~~ LCM***   * ***The proprietary/privacy information should not be disclosed*** * ***The performance, ~~generalization~~ and specification impact should be considered***   + ***~~Assistance information can be considered when there is some corresponding evaluation result(s).~~*** |
| Qualcomm | We believe the prospect of assistance information being “deemed” proprietary should not dissuade companies from evaluating the efficacy of assistance information in their simulations. Is there a clear definition or understanding of proprietary information? Is there a consensus among companies on what exactly is proprietary and what is not? We believe the sequence of events should be as follows: 1) Companies explore and evaluate the efficacy of assistance information in their simulations and then if performance gains are justified 2) debate about the feasibility of signaling assistance information from the viewpoint of revealing proprietary information, rather than the other way round. As mentioned before, there is a precedent to this in 3gpp. gNB beam shape information was “deemed” proprietary but in Rel-17 positioning indication of beam shape information from gNB to UE is already supported in Spec, due to performance gains observed in evaluations.  So, we suggest rewording the proposal to the following:  ***Conclusion 3.5.2:*** ***For the determination/selection of assistance information (if supported),***   * ***The performance and specification impact should be considered*** * ***The prospect of revealing proprietary/privacy information should be considered.*** |
| Xiaomi | We share same view as Qualcomm that the definition of proprietary/privacy information is not clear now, so it is better to evaluate the performance based on some assistance information first.  In addition, we share same view as Huawei that performance includes generalization, it is unnecessary to list it separately. |
| vivo | To answer questions on adding “generalization” (maybe repeat what we have clarified during online discussion):  Our understanding on generalization is it includes two aspects, where one is the performance for generalization, and the other is the scalability of applying a particular AI/ML model. One example for the second aspect is whether the output/input size a specific model can be scalable for multiple configurations/scenarios. These is no performance evaluation needed in this case. Hence to take into account the second aspect, we need to add either “generalization” or “scalability” into this bullet.  For the discussion on propriety/privacy information, we believe companies have different understanding on what information is propriety or privacy. Hence it is not that useful to give a hard limit for now. We can simply say that we need to consider propriety/privacy information protection.  Therefore, our suggestion will be  ***Conclusion 3.5.2b:*** ***For the determination/selection of assistance information (if supported), e.g., for data collection, AI/ML model inputs and other components of LCM***   * ***~~The proprietary/privacy information should not be disclosed~~*** * ***The performance, generalization, proprietary/privacy information protection and specification impact should be considered*** |
| Samsung | Share similar view with HW. |
| LGE | Slightly prefer the version from HW. Regarding the first bullet, we sympathize QC’s comment that ‘proprietary/privacy information’ is undefined and may be understood differently across different companies. But to our understanding, this is something high-level principle to keep proprietary/privacy information as much as possible as 3GPP has been done for decades. It may be revised as below to address QC’s concern:   * ***~~The proprietary/privacy information should not be disclosed~~ Strive not to disclose proprietary/privacy information*** |
| CAICT | We prefer HW’s wording. |
| NTT DOCOMO | Share the similar view with LGE, and we slightly prefer to discuss the feasibility of the proposed assistance information in terms of proprietary/privacy aspects before evaluating the efficiency of assistance information. It could reduce the unnecessary workload of exploring and evaluating the assistance information which is not feasible in the end due to proprietary/privacy issues. |
|  |  |
|  |  |
|  |  |
|  |  |

###### List of assistance info 3.5.3

As said before, many companies thought the contents of assistance information are not clear and request the proponents to disclose the detailed assistance information and how to use it. Meanwhile, since many discussions of spec impact are related to the detailed assistance information, it is difficult to make any progress there if the group cannot converge to some type(s) of assistance information.

Meanwhile, operators (e.g., DCM) encourage companies to share their views on what deployment information (e.g., beam information or Rx beam ID) can be exchanged between UE and gNB.

In order to have a clear picture on the potential candidates of assistance information, it would be beneficial to collect a list of the typical assistance information suggested by companies. Then, we can do further down-selection based on the list.

Companies are encouraged to provides inputs on the assistance information and their preference.

* The proponent(s) of a given type of assistance information is also encouraged to provide the reference (e.g., tdoc number) to contribution(s) that illustrates the detailed solution and corresponding evaluation results, so that other companies can double check and better understand the benefits.

|  |  |  |
| --- | --- | --- |
| Common assistance information for UE-side model and NW-side model | | |
| Assistance information | Support | Not support |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
| Specific assistance information for NW-side model | | |
| Assistance information | Support | Not support |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
| Specific assistance information for UE-side model | | |
| Assistance information | Support | Not support |
| NW-side beam shape information (3dB beamwidth, beam boresight directions, beam shape, etc.) | Qualcomm |  |
| Beam angle information in 2D/3D (e.g. 2D/3D coordinate value) | LGE |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

Companies can provide detailed inputs (if any) in the following table.

|  |  |
| --- | --- |
| Company | Comments |
| vivo | **For NW-side model:** Expected Rx beam ID/angle, Rx beam/Beam pair ID/angle  **For UE-side model:** Expected Tx beam ID/angle, Tx beam/Beam pair ID/angle  Details can be found in R1-2208636 and R1-2208637 |
| MediaTek | Beam ID and beam angle, Beam pair ID and beam angles |
| Qualcomm | The efficacy of indicating information about gNB beam shape has been indicated in our Tdoc submission for evaluations agenda item. This is mentioned as an example, there needs to be further evaluations to study the performance gains enabled by assistance information. |
| Xiaomi | **For UE-side model**: at least the Tx beam ID of gNB (and it can be indicated by RS ID implicitly)  **For NW-side model**: at least the Rx beam ID of UE |
| LGE | Beam angle information is essential for training/inference. Updated table above. |
| NEC | For NW-sided model: Rx beam ID/angle |

###### Prioritization of alternatives 3.5.4

Many contributions suggest to make some down-selection on the alternatives for AI/ML inputs. The following two tables summarize the views of the contributions submitted to this meeting and the inputs in the FL summaries of the last meeting.

|  |  |  |
| --- | --- | --- |
| BM-Case1 | | |
|  | Prioritize | Down-prioritize |
| Alt.1: Only L1-RSRP measurement based on Set B | Huawei[2], ZTE[3],CATT[11], NVIDIA[26], Spreadtrum[4], Nokia[20], Fujitsu, NEC, MTK, NVIDIA | IDC[6] |
| Alt.2: L1-RSRP measurement based on Set B and assistance information | vivo[5], LGE[9], CATT[11], Nokia[20], vivo, NEC | Huawei[2]? ZTE[3], Spreadtrum[4],Google, |
| Alt.3: CIR based on Set B | IDC[6] (only for FR1), Google[8], Sony[14], | Huawei[2], CATT[11], ZTE[3], Spreadtrum[4], Xiaomi[18] |
| Alt.4: L1-RSRP measurement based on Set B and the corresponding DL Tx and/or Rx beam ID | ZTE[3], IDC[6], CATT[11], NVIDIA[26] ? , Spreadtrum[4], Fujitsu,vivo,NEC, NVIDIA |  |

|  |  |  |
| --- | --- | --- |
| BM-Case2 | | |
|  | Prioritize | Down-prioritize |
| Alt.1: Only L1-RSRP measurement based on Set B | Huawei[2], ZTE[3],CATT[11], NVIDIA[26], Spreadtrum[4], Nokia[20], Fujitsu,NEC, MTK, NVIDIA | IDC[6] |
| Alt 2: L1-RSRP measurement based on Set B and assistance information | vivo[5], LGE[9], CATT[11], Nokia[20], vivo,NEC, | Huawei[2] ZTE[3], Spreadtrum[4], |
| Alt.3: L1-RSRP measurement based on Set B and the corresponding DL Tx and/or Rx beam ID | ZTE[3], IDC[6], CATT[11], NVIDIA[26] ? , Spreadtrum[4], Fujitsu, vivo, NEC, NVIDIA |  |

Google[8] also proposes some new alternatives for Case1 and Case2, for example, CIR plus L1-SINR for the input of Case1, CIR and CIR plus L1-SINR for the input of Case2.

Some companies suggeste that the alternatives will be different depending whether it is UE-side model or NW-side model.

Moderator feels that if there is any progress for “clarification 3.5.1” and/or “List of assistance info 3.5.3”, some companies may change their views on these alternatives. Thus, for this topic, moderator suggests to continue collecting companies’ view on the above issues. Some proposal may be suggested based on the inputs and progress of other topics.

|  |  |
| --- | --- |
| Company | Comments |
| Google | Regarding the prioritization and de-prioritization, we think it is better to check the results before we make final decision. |
| Vivo | Better to study and evaluate the alternatives in EVM first. |
| Sony | For BM-Case 1, we still support Alt 3. |
| MediaTek | L1-RSRP should be the baseline. CIR and other information can be studied on top of that. |
| HW/HiSi | I removed the question mark behind Huawei to down-prioritize Al2 for BM-Case 2. |
| Qualcomm | We believe CIR and other inputs should not be precluded at this point. Future evaluation results can be the criterion for potential future down-selection. |
| Xiaomi | L1-RSRP can be the baseline. In addition, L1-RSRP+beam ID can be prioritized. |
| LGE | Unless we intend to specify AI/ML model input parameters in 3GPP specification, we would not need down-selection. |
| ZTE | Focusing the AI input on measured RSRP and/or beam ID would be a good starting point, in which case the standardization workload and AI model complexity would be relatively low. |

## Output of BM-Case1 and BM-Case2

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

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

The related proposals/observations are copied as below:

|  |  |
| --- | --- |
| FUTUREWEI[1] | *Observation 3: Model outputs are typically used internally so unless there are standards impacts involved, they don’t need to be explicitly specified in the standards.*  *Proposal 3: Specify model outputs only when standards impact is involved while companies are encouraged to share their model output for AI/ML based beam management.* |
| Huawei[2] | *Observation 2: For the alternatives for AI/ML output for BM-Case 1 and BM-Case 2, Alt. 2 (beam ID and other information) has too many sub-options and for its further study a down-selection within Alt.2 is necessary. Alt. 3 (beam angle and RSRP) can be seen as a further sub-option of Alt.2.*  *Proposal 5: For BM-Case1 and BM-Case2, consider Alt. 1 as the baseline for the assumption on the AI/ML model output:*   * *Alt.1: Tx and/or Rx Beam ID(s) and/or the predicted L1-RSRP of the N predicted DL Tx and/or Rx beams*    + *E.g., N predicted beams can be the Top-N predicted beams* |
| ZTE[3] | *Observation 4: Alt.1 can provide better standard compatibility and less additional standardization effort since only beam ID and the predicted RSRP are considered as the AI output.*  *Proposal 6: Focus the discussion on Alt.1 as a starting point. Alt.3 can be postponed until the relationship between output beam angle and TCI state is clear.*  *Proposal 8: For temporal domain beam prediction, focusing the AI input and output on measured RSRP and/or beam ID would be a good starting point, in which case the standardization workload and AI model complexity would be relatively low.* |
| Vivo[5] | *Proposal 3: Support to prioritize following AI output for further study on specification impact:*   * *Tx and/or Rx Beam ID(s)/angle(s) and/or the predicted L1-RSRP of the N predicted DL Tx and/or Rx beams.*   + - * + *The N predicted Tx/Rx beams can be produced according to the expected beam information input to the AI model*         + *FFS: study global beam ID or local beam ID*         + *FFS: study global beam information, e.g. global beam ID or beam angle, with minimum exposures of implementation details*   *Proposal 4: Suggest to deprioritize Alt.2, i.e. Tx and/or Rx Beam ID(s) of the N predicted DL Tx and/or Rx beams and other information, for further study specification impact.* |
| IDC[6] | *Proposal 9: Support ‘Tx and/or Rx Beam ID(s) and/or the predicted L1-RSRP of the N predicted DL Tx and/or Rx beams’ as a baseline.*  *Proposal 10: ‘Tx and/or Rx Beam ID(s) of the N predicted DL Tx and/or Rx beams and other information’ can be considered with LOS probability.*  *Proposal 11: Benefits from utilization of TX/Rx beam angles should be clarified.* |
| OPPO[7] | *Proposal 6: For the output of AI/ML model for BM-Case1 and BM-Case2, suggest to include at least*   * + *Tx and/or Rx Beam ID(s)*   + *The predicted L1-RSRP of the predicted Top-K DL Tx and/or Rx beams*   + *Note: the above output should be extended for F time instances for BM-Case2* |
| Google[8] | *Proposal 3: For spatial domain beam prediction, support the best beam possibility for each beam in Set A as the output.*  *Proposal 4: When AI/ML model is implemented in the NW side, the output for the AI/ML for spatial domain beam prediction with spec impact should be the reference angle for DL Rx beam refinement (Alt3).*  *Proposal 5: When AI/ML model is implemented in the UE side, the output for the AI/ML model for spatial domain beam prediction with spec impact should be the reference angle for DL Tx beam refinement (Alt3).*  *Proposal 9: For time-domain beam prediction, support the best beam possibility for each beam in Set A as the output.*  *Proposal 10: When AI/ML model is implemented in the NW side, the output for the AI/ML for time domain beam prediction with spec impact should be the reference angle for DL Rx beam refinement (Alt3).*  *Proposal 11: When AI/ML model is implemented in the UE side, the output for the AI/ML model for time domain beam prediction with spec impact should be the reference angle for DL Tx beam refinement (Alt3).* |
| Ericsson[10] | The exact AI/ML model output is not expected to be standardized; however, the potential AI/ML output (after post-processing) might be subject for standard impact.  *Proposal 4 Further define the FFS on AI/ML output after sufficient progress is made on studying the specification impact for AI/ML model inference aspects* |
| Intel[13] | *Proposal 3: For BM-Case1 and 2, Alt-1 should be considered as the baseline use case, with potential specification impact on how beam IDs are mapped in the spatial domain.* |
| Sony[14] | *Observation 1: Different AI/ML models are trained based on different objective functions. The output TX/RX beam ID may be chosen based on different criteria.*  *Proposal 1: In output of AI/ML, should clearly indicate the criterion associated with the predicted beam ID in BM-case1 and BM-case2 for example TX beam ID for maximum dwelling time, TX/RX beam ID for maximum RSRP, etc.* |
| Lenovo[15] | *Proposal 3: Tx and/or Rx Beam ID(s) and the predicted L1-RSRP of the N predicted DL Tx and/or Rx beams should be taken as the AI model output at least for UE-centric AI inference.*  *Proposal 4: When specify the AI model output, we should consider that it may be used for model monitoring.* |
| NEC[16] | *Proposal 1: Support selecting Top-N1 DL Tx and/or Rx beams according to some pre-defined rules, e.g., a sum probability of being the best beam higher than a threshold, L1-RSRP higher than a threshold.* |
| CIACT[17] | *Proposal 4: Tx and/or Rx Beam ID(s) and/or the predicted L1-RSRP of the N predicted DL Tx and/or Rx beams should be used as baseline for further comparison for BM-Case1 and BM-Case2.* |
| Nokia[20] | *Observation 13: For the sub-use case BM-Case1, depending on the ML model used, the model may output other information (e.g., a QoS-based metric of Tx beams) which are useful to determine Top-N1 beams or to report additional parameters other than to Top-N1 beams.*  *Proposal 5: Regarding the sub-use case BM-Case1 and BM-Case 2, further study the following alternatives for AI/ML output:*   * *Alt.1 : Tx and/or Rx Beam ID(s) and/or the predicted L1-RSRP of the N predicted DL Tx and/or Rx beams* * *Alt.2: Tx and/or Rx Beam ID(s) of the N predicted DL Tx and/or Rx beams and other information*   + *FFS: other information (e.g., QoS-based metric, probability for the beam to be the best beam, the associated confidence, beam application time/dwelling time, Predicted Beam failure)* * *Alt.3: Tx and/or Rx Beam angle(s) and/or the predicted L1-RSRP of the N predicted DL Tx and/or Rx beams* |
| MTK[23] | *Proposal 5: RAN1 will discuss and define the method(s) to select the top N DL Tx and/or Rx beams.* |
| Samsung[27] | *Proposal 7: For AI/ML output for beam prediction, Alt 1 (beam ID + L1-RSRP) is preferred.* |

###### List of other info 3.6.1

Many companies think the contents of other information (in Alt.2) are not clear and request the proponents to disclose more details of other information. In the contribution, some companies propose some types of the other information as below.

|  |  |
| --- | --- |
| Other information for Alt.2 | |
| IDC[6] | LOS/NLOS possibility |
| Google[8] | The possibility for each beam to be the best beam |
| Nokia[20] | QoS-based meric |
| Some other companies | Confidence of the predicted beam |

In order to have a clear picture on Alt.2, it would be beneficial to collect a list of the typical types of other information suggested by companies. Then, we can do further down-selection based on the list if Alt.2 is supported.

Companies are encouraged to provides inputs on the other information and their preference.

|  |  |  |
| --- | --- | --- |
| Other information | Support | Not support |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

The proponent(s) of a given type of other information is encouraged to explain the benefits and whether this type of other information has any spec impact.

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

###### Clarification 3.6.2

Regarding Alt.3, many companies raise the concern how to define and get the beam angles, e.g.,

* Some companies think that the beam information about the NW is proprietary and should not be disclosed
* Some companies think there is no fixed reference for the determination of beam angles.

In summary, many companies doubt the feasibility of beam angle acquisition in Alt.3, or Alt.3 is a special case of Alt.1.

Moderate suggests the proponent of Alt.3 to clarify the definition and acquisition of beam angle, and to address the concerns from other companies.

|  |  |
| --- | --- |
| Company | Comments |
| Google | Beam angle can be defined as target/reference ZoD/AoD for Tx beam prediction and target/reference ZoA/AoA for Rx beam prediction |
|  |  |
|  |  |
|  |  |

###### Prioritization of alternatives 3.6.3

Many contributions suggest to make some down-selection on the alternatives for AI/ML output. The following table summarizes the views of the contributions.

|  |  |  |
| --- | --- | --- |
| AI/ML output for BM-Case 1 and BM-Case 2 | | |
|  | Prioritize | Down-prioritize or postpone |
| Alt.1: Tx and/or Rx Beam ID(s) and/or the predicted L1-RSRP of the N predicted DL Tx and/or Rx beams | Huawei[2], ZTE[3], vivo[5], IDC[6], OPPO[7], Intel[13], Lenovo[15], CIACT[17], Samsung[27], |  |
| Alt.2: Tx and/or Rx Beam ID(s) of the N predicted DL Tx and/or Rx beams and other information | IDC[6],Google[8], | vivo[5], |
| Alt.3: Tx and/or Rx Beam angle(s) and/or the predicted L1-RSRP of the N predicted DL Tx and/or Rx beams | Spreadtrum[4], vivo[5],  Google[8] | ZTE[3], |

Moderator feels that if there is any progress for “List of other info 3.6.1” and/or “Clarification 3.6.2”, some companies may change their views on these alternatives. Thus, for this topic, moderator suggests to continue collecting companies’ views. Some proposal may be suggested based on the inputs and progress of other topics.

|  |  |
| --- | --- |
| Company | Comments |
| Sony | In alternative 2, the “other information”, this is needed to define what the output of AI is for, or the selection criteria of the beam. For example, this information can be the maximum dwelling time of the predicted beam ID, or the L1-RSRP higher than a threshold, a sum probability of being the best beams higher than a threshold, RSRP corresponding to the expected Tx and/or Rx beam direction(s). |
| Xiaomi | We prefer to prioritize Alt 1. Since the other information is not clear in Alt 2 and the beam angle is not clear in Alt 3. |
| LGE | What is the implication of down-selecting AI/ML model output in 3GPP? Would this restrict AI/ML implementation to a specific way? |
| ZTE | Focus the discussion on Alt.1 as a starting point. In Alt.2, specific definition of the other information is needed with further evaluation results being provided by the proponents. In Alt.3, the beam angle is included as part of the model output, which is related to the beam implementation issue and hard to be aligned among companies. Besides, in the current specification, beam indication is performed through the configured TCI state. It is also unclear how to establish a corresponding relationship between the output beam angle and the TCI state. |

## Other use cases

In RAN1#109e meeting, sub use cases and categories were summarized as below:

|  |  |
| --- | --- |
| Category | Sub use case |
| Cat1:  Spatial-domain DL beam prediction | **BM-Case1:** Spatial-domain DL beam prediction for Set A of beams based on measurement results of Set B of beams |
| **BM-Case3:** Beam prediction for higher frequency band (e.g., a band in FR2) based on measurement results of lower frequency band(s) (e.g., a band in FR1) |
| **BM-Case4:** Beam prediction based on UE positioning/trajectory |
| **BM-Case6:** Spatial-domain UL beam prediction for Set A of beams based on measurement results of Set B of beams |
| **BM-Case9:** Joint DL/UL beam pair link prediction |
| Cat2:  Time-domain DL beam prediction | **BM-Case2:** Temporal DL beam prediction for Set A of beams based on the historic measurement results of Set B of beams |
| Cat3: Others | **BM-Case7:** beam measurement feedback compression |
| **BM-Case8:** The beam-specific parameter optimization |

There are some discussions on these sub use cases in the tdocs. The related proposals/ observations are copied as below:

|  |  |
| --- | --- |
| ZTE[3] | *Proposal 2: Since the time unit of this study item is limited, we suggest to focus on the sub use cases BM-Case1 and BM-Case2 in Rel-18, and other sub use cases can be discussed until solid AI framework has been made in Rel-18.* |
| Vivo[5] | *Proposal 7: No need to support other sub use cases in addition to BM-Case1 and BM-Case2 in Rel-18.* |
| IDC[6] | *Proposal 4: 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.* |
| OPPO[7] | *Proposal 7: Study only BM-Case1 and BM-Case2 as representative use case with high priority.* |
| LGE[9] | *Proposal #6: BM sub use cases other than BM-Case1 and BM-Case2 are deprioritized during this SI.* |
| CATT[11] | *Observation 1: BM-Case3 is already included in BM-Case1 and BM-Case2 for Alt.1, i.e., Set A and Set B are different (Set B is NOT a subset of Set A).*  *Observation 2: The UE positioning/trajectory information can be as assistance information of AI/ML model inputs for beam prediction, which can be studied in BM-Case1 and BM-Case2.*  *Observation 3: There is no spec impact on spatial-domain UL beam prediction for Set A of beams based on measurement results of Set B of beams.*  *Observation 4: The beam measurement feedback compression is similar with the use case of CSI feedback compression.*  *Observation 5: Parameter optimization to improve performance of multi-beam system, e.g., beam-based mobility enhancement, is important but has higher complexity.*  *Observation 6: The spec impact of joint DL/UL beam pair link prediction is the same with BM-Case1 and BM-Case2.*  *Proposal 3: For AI/ML-based beam management, BM-Case3 and BM-Case4 can be studied together with BM-Case1 and BM-Case2. Thus, there is no need to specifically study BM-Case3 and BM-Case4.*  *Proposal 4: For AI/ML-based beam management, the following sub use cases are deprioritized:*   * *BM-Case6: Spatial-domain UL beam prediction for Set A of beams based on measurement results of Set B of beams;* * *BM-Case7: beam measurement feedback compression;* * *BM-Case8: Parameter optimization to improve performance of multi-beam system;* * *BM-Case9: Joint DL/UL beam pair link prediction.* |
| Sony[14] | *Observation 2 : Beam prediction in mmWave can be assisted by CSI information at low frequency.*  *Proposal 2 : Support BM-case3: Beam prediction for higher frequency band (e.g., a band in FR2) based on measurement results of lower frequency band(s) (e.g., a band in FR1).* |
| Lenovo[15] | *Proposal 5: Beam prediction at gNB/TRP side with model management-related collaboration between gNB and UE can be taken as a sub-use case for beam management in predictable trajectory scenario.* |
| Xiaomi[18] | *Proposal 1: For AI/ML-based beam management, only support BM-Case1 and BM-Case2* |
| NVIDIA[26] | *Proposal 1: Beam prediction in spatial domain and beam prediction in time domain should be the focal point for studying AI/ML based algorithms for beam management.* |
| DCM[28] | *Proposal 1: Prioritize the discussion of spatial-domain DL beam prediction and temporal DL beam prediction from other sub use case.* |
| KT[31] | *Proposal 1: Study BM-Case1 and BM-Case2 as representative sub use cases.* |

###### (Closed) Conclusion 3.7

Based on the contribution, companies’ view on the other sub use cases are summarized as below

|  |  |
| --- | --- |
|  | Supporting companies |
| BM-Case3 | Sony[14], IDC[6], CATT[11] (be part of Case1/2), |
| BM-Case4 | CATT[11] (be part of Case1/2), Lenovo[15] |
| BM-Case6 | Samsung[27] |
| BM-Case7 |  |
| BM-Case8 |  |
| BM-Case9 |  |
| Deprioritize all other sub use cases | ZTE[3], vivo[5],OPPO[7],LGE[9], xiaomi[18], NVIDIA[16],DCM[28], KT[31], |
| Deprioritize BM-Case 6/7/8/9 | CATT[11], |

Considering there is a deadline to determine the final representative sub use cases before or in the Dec. RAN plenary meeting, there is only one remaining meeting (i.e., Nov. RAN1 meeting) for RAN1 evaluation. Thus, the group have to make the final decision on whether any additional sub use case is support or not in this meeting.

From the above table, we can see that the views on whether to support other sub use cases or not are quite diverging. Case3 gets more supporters compared to other cases. Meanwhile, in the last meeting,

* BM-Case3 got 6 supporting companies (MTK, Google, Sony, Apple, IDC, Fujitsu). Other sub use cases get fewer supporting companies.
* 15 companies (LGE, ZTE, NEC, CAICT, NVIDIA, FUTUREWEI, Xiaomi, Spread, vivo, QC, Fujitsu, HW, DCM, SS, CMCC) suggested to deprioritize all other sub use cases.

According to the submitted contributions, the situation seems unchanged on the support of additional sub use cases. Thus, a conclusion as below seems the only possible way.

***Conclusion 3.7:***

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

|  |  |
| --- | --- |
| Company | Comments |
| Google | It seems BM case3/4/6/9 are already agreed as part of BM case 1? Maybe this conclusion is not needed.  Mod: In moderator’s understanding, they are not included in BM-Case1. |
| Nokia | OK. The wording could be revised, “support on studying”. |
| NEC | Fine |
| Spreadtrum | Support |
| LGE | Support. |
| Ericsson | Support |
| CATT | Just for clarification, is BM-Case3 already included in BM-Case1 and BM-Case2 for Alt.1, i.e., Set A and Set B are different (Set B is NOT a subset of Set A)?  For other use cases, i.e., BM-Case 4/6/7/8/9, we agree to not support in R18.  Mod: In moderator’s understanding, it is not included |
| NTT DOCOMO | Support. |
| vivo | Ok for the above conclusion. |
| HW/HiSi | Support |
| Sony | We support BM-Case 3 as a part of BM-Case 1/ 2 to be further studied.  Mod: In moderator’s understanding, BM-Case3 is not included in BM-Case1/2 so far. |
| Samsung | Ok with the proposal in general. |
| Xiaomi | Support |
| Lenovo | Support |
| NVIDIA | Support |
| ZTE | Support |

###### Conclusion 3.7b

In addition to the discussions in GTW, some clarifications from moderation perspective were made for the concerns raised in GTW discussion:

* + - * Regarding the note “For BM-Case1 and BM-Case2, Beams in Set A and Set B can be in the same Frequency Range”, the background is as below
        + The original version is that Set A and Set B are in the same band.
        + In the GTW session, Apple (i.e., Yushu) commented that RAN4 spec supports the common beam for multiple bands within the same FR and thereby should not restrict Set A and Set B in the same band. Thus, it was changed to be “can be in the same Frequency Range” just in order to avoid the restriction “in the same band”
      * Regarding the concern the conclusion will impact the discussion on Alt.3: CIR based on Set B, my understanding is that there are separate issues.

In order to address the concern about the impact of conclusion 3.7/3.7a on the AI/ML model inputs (e.g., Alt.3: CIR based on Set B), a note is added as the sub-bullet. Let’s check whether it is acceptable to all companies.

***Conclusion 3.7b:***

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

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

|  |  |
| --- | --- |
| Company | Comments |
| NVIDIA | Support |
| HW/HiSi | Support |
| xiaomi | Support |
| vivo | Generally okay. The following is suggested to make the sub-bullet more clear.  ***Conclusion 3.7b:***  ***For AI/ML based beam management, RAN1 has no consensus to support on studying any other sub use case in addition to BM-Case1 and BM-Case2***   * ***Note: this conclusion is independent of the discussion on the alternatives of AI/ML model inputs for the agreed use cases***   Mod: Updated the proposal by adding “for BM-Case1 and BM-Case2” |
| LGE | OK with either previous version or revised version. |
| ZTE | Support |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

# Spec impact

## General views

The related proposals/ observations are copied as below:

|  |  |
| --- | --- |
| vivo[5] | *Proposal 8: For case 1 and case 2 of beam management, both collaboration level level-y, and collaboration level-z can be considered.* |
| Google[8] | *Proposal 14: For AI/ML based BM, the study should be based on both Rel-17 unified TCI framework and Rel-15/Rel-16 BM framework.*  *Proposal 15: The study of AI/ML based BM should consider both FR1 and FR2.* |
| Intel[13] | *Observation 3: One possible area of specification impact for AI/ML model integration may be for triggering of beam measurement reports and reference signal transmissions, as well new L1 reporting formats.* |

**Mod recommendation:** TBD

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

## Life cycle management

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

|  |
| --- |
| **RAN1#110**  Agreement  Study the following aspects, including the definition of components (if needed) and necessity, in Life Cycle Management   * Data collection   + Note: This also includes associated assistance information, if applicable. * Model training * [Model registration] * Model deployment   + Note: Terminology is to be defined. ~~This includes process of compiling a trained AI/ML model and packaging it into an executable format and delivering to a target device.~~ * [Model configuration] * Model inference operation * Model selection, activation, deactivation, switching, and fallback operation   + ~~Note: some of them to be refined~~ * Model monitoring * Model update   + Note: Terminology is to be defined. This includes model finetuning, retraining, and re-development via online/offline training. * Model transfer * UE capability   Note: Some aspects in the list may not have specification impact.  Note: Aspects with square brackets are tentative ~~and pending terminology definition~~.  Note: More aspects may be added as study progresses. |

The related proposals/ observations are copied as below:

|  |  |
| --- | --- |
| FUTUREWEI[1] | *Proposal 6: Regarding AI/ML-based beam management, study the standards impact related to AI/ML model selection/configuration (like activation/deactivation) in case multiple trained AI/ML models are deployed.*  *Proposal 8: Study Standards impact related to supporting model generalization across scenarios and/or configurations.* |
| Huawei[2] | *Proposal 7: For the study of life cycle management for beam management use case, discuss use case specific procedures in 9.2.3.2, including training, updating, deployment, data collection, inference, monitoring, fallback, and UE capability.*   * *FFS: [model registration], and [model configuration]* |
| vivo[5] | *Proposal 9: Take the following supportable model update choices as one aspect for defining model update levels of beam management.*   * *Choice 0: No model update during lifecycle management* * *Choice 1: Updating model parameter or structure w/o model transfer* * *Choice 2: Updating model parameter or structure with model transfer* * *Study the lifecycle management signaling and procedures for each of the collaboration levels and model updating choices.*   *Proposal 10: At least the following life cycle management component need to be studied for beam management: model activation, data collection for model inference, model inference, data collection for model monitoring, model monitoring and model deactivation.*  *Proposal 17: Study signaling aspects enhancement related to the procedure of model transfer, model confirmation and model activation, if AI/ML model training at NW side and AI/ML model inference at UE side.* |
| OPPO[7] | *Proposal 15: For BM-Case1 and BM-Case2, study mechanisms to enable generalization of AI/ML model under heterogeneous environments and beam-related configurations.* |
| Ericsson[10] | *Proposal 17 Study mechanisms to activate/deactivate beam prediction AI/ML models, and potential fallback mechanisms* |
| CATT[11] | *Proposal 11: Regarding BM-Case1 and BM-Case2, if the model is inferenced at UE side, for collaboration Level y, the following aspects can be studied as model registration information which UE should provide to gNB:*   * *Model ID;* * *Model functionality, e.g., BM-Case1/BM-Case2 or DL beam pair/Tx beam/Rx beam prediction;* * *Information of model inputs, e.g., the number of DL Tx beams or beam pairs in Set B;* * *Information of model outputs, e.g., the number of predicted beam.* |
| Fujitsu[12] | *Proposal 6: For AI model life cycle management of BM-Case1 and BM-Case2, support to investigate the necessity and/or specification impacts from the following aspects*   * *Data collection* * *Model training* * *Model registration* * *Model deployment* * *Model configuration* * *Model inference operation* * *Model selection, activation, deactivation, switching, and fallback operation* * *Model monitoring* * *Model update* * *Model transfer* * *UE capability*   *Proposal 7: Study the potential specification impacts on model selection for DL beam prediction on AI/ML from the following aspects*   * *Mechanism to facilitate the management on multiple models* * *New signaling/procedure on model selection* |
| Lenovo[15] | *Proposal 11: Dynamic switching between AI/ML based beam prediction and non-AI/ML based beam report should be supported* |
| NEC[16] | *Observation 1: For a sub use case, multiple AI/ML models may be arranged.*  *Proposal 4: Study the mechanism of model selection.* |
| CIACT[17] | *Proposal 6: If AI model transfer from NW side to UE side is supported, AI model transfer over air interface should be specified.*  *Proposal 7: NW side could send assistant information to UE side for AI model update.* |
| Xiaomi[18] | *Proposal 6: To discuss whether a common AI model or separate AI models will be trained for UE with different number of Rx beam.*  *Proposal 12: Study the mechanism for AI model update/disable/deactivation request from UE.*  *Proposal 13: Study the mechanism for AI model disable/deactivation/update by gNB.* |
| Nokia[20] | *Observation 14: In an online/continual learning scenario, as well as supervised learning, when the ML model selects suboptimal beams in terms of signal quality (e.g., due to insufficient ML model accuracy or to explore the action space), a fallback mechanism should be in place to guarantee successful data transmission. For instance, by configuring a known good beam as anchoring beam which can be used when ML selected beam fails.*  *Proposal 6: For online/continual learning-based beam prediction, further study fallback mechanisms in cases where the ML model selected beam fails.*  *Observation 25: NW may perform ML model monitoring to track variation of the ML model performance for all the served UEs. Also, the NW may be better aware of the propagation conditions for all the Ues in the sector area and indicate to the UE when it requires switching to a different ML model.*  *Proposal 20: RAN1 to study NW input to support ML model switching/(de)activation at UE for DL Tx beam or DL Tx-Rx beam pair prediction.* |
| Apple[24] | *Observation 1: AI/ML models can be crafted as a universal channel parameter estimator with good generalization or as a beam management database and associated query mechanism customized for a specific cell which is not expected to generalize well.*  *Proposal 1:*  *• For Model training at the NW side & inference at the NW side, study efficient signalling of set B selection or beam selection and RSRP representation.*  *• For Model training at the NW side & inference at the UE side, or model training at the UE side & model inference at the NW side, study model generalization performance, study model transfer/model delivery for cell-specific AI models and non cell-specific AI models.*  *• For Model training at the UE side, and inference at the UE side, study cell-specific signals to facilitate data collection.* |
| NVIDIA[26] | *Proposal 9: For AI/ML based beam prediction in spatial/time domain, study potential specification impact related to assistance signalling and procedure for model configuration, model activation/deactivation, model recovery/termination, and model selection.*  *Proposal 10: For AI/ML based beam prediction in spatial/time domain, study potential specification impact related to assistance signalling and procedure for model performance monitoring and model update/tuning.* |

###### Mod Recommendation 4.2.1

According to the achieved agreement(s), AI/ML model life cycle management (LCM) consists all the components related to the different stages/procedures of AI/ML operations. By reading the contributions, moderator feels that the components can be categorized as below:

* Cat1: There is no specific spec impact(s) mentioned by companies for AI/ML based beam management. For these components, moderator feels it is better to discuss the common design for all sub use cases in Agenda item 9.2.1. If some BM-specific spec impact(s) is identified later, a separate section will be added for the discussion.
* Cat2: There are some specific spec impacts mentioned by companies for AI/ML based beam management. For these components, a separate section is preserved for each of them to discuss the detailed spec impacts.

**Mod recommendation:** A separate section is used for each of the following LCM components, with the focus on BM-specific aspects:

* Data collection
* Model inference operation
* Model monitoring
* UE capability
* Note: separate section(s) will be added for other component(s) once some specific spec impact(s) for BM is identified.

|  |  |
| --- | --- |
| Company | Comments |
| Google | Support |
| Sony | What will be the difference with the 9.2.1 discussion?  Mod: In this agenda, we will focus on BM-specific aspects, whereas the common aspects/parts will be discussed in agenda 9.2.1 |
| Fujitsu | Support |
| MEDIATEK | Support |
| Qualcomm | We believe the discussions here should wait until the general aspects is elaborated and then BM-specific aspects can be discussed here.  Mod: In this agenda we only focus MB-specific aspect. If there is some potential conflicting/overlapping with agenda 9.2.1, companies can point out the detailed issue(s)/topic(s). There is no reason to hold on all discussion in each agenda to wait for the progress of general part. In my understanding, all the proposals in Section 4.3/4.4/4.5 seems agnostic to the design of agenda 9.2.1. |
| Xiaomi | It is better to revise “Data collection” to “Data collection for model training”  Mod: It does not matter since it is only to share the consideration on how to organize the discussion, not an “formal proposal”. Moreover, these terminologies are just copied from the agreement |
| Samsung | Ok with the recommendation. |
| NEC | Agree |
| CAICT | Support |
| NTT DOCOMO | Support the proposal. |
| ZTE | Support |

## Data collection

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

|  |
| --- |
| **RAN1#110**  Agreement  For the data collection for AI/ML model training (if supported), study the following aspects as a starting point for potential necessary specification impact:   * Signaling/configuration/measurement/report for data collection, e.g., signaling aspects related to assistance information (if supported), Reference signals * Content/type of the collected data * Other aspect(s) is not precluded |

The related proposals/ observations related are copied as below:

|  |  |
| --- | --- |
| FUTUREWEI[1] | Proposal 5: In AI/ML-based beam management, when model training is at NW side and “Alt.3: Beam pair prediction” is adopted, study the standards impact to enable gNB to map the received L1-RSRP measurements to the corresponding Tx-Rx beam pairs. |
| Huawei[2] | Proposal 10: RAN1 to further study the potential spec impact of data collection from a realistic network for training from the following aspects:   * For reference signal, enhanced RS design can be considered, e.g., RS design for AI/ML specific RSRP measurement and enhancement of RS for improving data sample accuracy * For UE measurement/report, new RSRP and/or CRI/SSBRI report behavior can be considered * For the signaling/configuration, signaling to trigger/configure/request data collection window can be considered |
| ZTE[3] | Proposal 14: Further research on enhanced signaling mechanisms and auxiliary information transmission is needed to enable training data collection. |
| Vivo[5] | Observation 5: Study the signaling aspects related to Rx beam or Rx beam pattern indication in training stage for Alt.1.  Observation 6: Study report enhancement as well as assistance information report for AI model training purpose at gNB side.  Observation 7: A training request signaling may be needed for enhanced beam pair prediction and (enhanced) DL Tx beam prediction with Alt.2 in training stage.  Observation 8: P1/P2 resource related information request, at least including minimum number of Tx beams and Tx beam pattern, should be used with specification impact.  Observation 9: Tx beam information and/or expected beam information as assistance information should be noticed to UE in advance or with resource configuration.  Observation 10: DL Rx beam prediction with Alt.2 may have some specification impacts on training request signaling, P3 resource related information request, and Tx beam information indication.  Observation 11: In Alt.3 training stage, AI model should be trained in network side with considering to adapt various UE capabilities.  Observation 12: Similar specification impacts can be obtained from Alt.3 and Alt.1 with same beam prediction solutions with generalization consideration.  Proposal 11: For AI/ML model training at NW side, at least study the following aspects for potential necessary specification impact:   * Signaling aspects related to Rx beam or Rx beam pattern indication/configuration * Report enhancement to carry more L1-RSRP and/or Rx beam information   Proposal 12: For AI/ML model training at UE side, at least study the following aspects for potential necessary specification impact:   * P1/P2 training request signaling to gNB * P1/P2 resource related information request to gNB , at least including Tx beam pattern, minimum number of Tx beams * Further study if P3 is proven beneficial * P3 training request signaling to gNB * P3 resource related information request to gNB, at least including Tx beam pattern, minimum number of Tx beam repetitions |
| OPPO[7] | Proposal 8: Study data collection for AI/ML model training with legacy beam measurement and reporting as a starting point. |
| LGE[9] | Proposal #1: For the UE AI/ML input, Alt2 can be considered. For the assist information for input, output, training, and inference, consider to express Set A and Set B beams on a pre-defined or configured beam grid. |
| Ericsson[10] | Proposal 5 The necessity of collecting assistance information together with the radio-measurement should be firstly studied and justified before discussing the method for collecting this type of data.  Proposal 6: Based on the study outcome from RAN1, study the following aspects for data collection for the beam management use case in RAN2:   * 1. DL-RS or UL-RS resource set configuration,   2. signaling for collected assistance information, if justified   3. signaling and configurations to support UE performing data logging/collection for model training,   4. signaling and configurations to support UE reporting the collected/logged data to the NW,   5. signaling for indicating UE capability for data collection.   Proposal 7 Consider mechanisms for reducing the radio resource overhead, memory consumption and power consumption for data collection  Proposal 8 Study mechanisms for improving the quality of the collected data for the considered beam prediction use cases. |
| CATT[11] | Proposal 5: Regarding the data collection for training/fine-tuning/update in BM-Case1 and BM-Case2,   * For Alt.1, gNB needs to send RS in both Set A and Set B to UE; * For Alt.2, gNB needs to send RS in Set A and informs the beam pattern of Set B to UE; * For Alt.3, gNB needs to send RS in Set A (i.e., Set B) to UE.   Proposal 6: Regarding the data collection for BM-Case1 and BM-Case2, if training/fine-tuning/update is performed at gNB side, the UE needs to report the measurement results (e.g., L1-RSRP) of Set B as model inputs and Top-N beam ID of Set A as the label of model outputs to gNB. Whether beam ID or other assistance information is needed as model inputs should be further studied.  Proposal 7: Regarding the data collection for BM-Case1 and BM-Case2, if training/fine-tuning/update is performed at UE side, the UE needs to get the measurement results (e.g., L1-RSRP) of Set B as model inputs and Top-N beam ID of Set A as the label of model outputs. Whether beam ID or other assistance information is needed as model inputs should be further studied.  Proposal 8: Regarding the data collection for inference in BM-Case1 and BM-Case2, gNB needs to send RS in Set B to UE. |
| Fujitsu[12] | Observation 1: UCI reporting overhead is increased a lot for DL beam predication on AI/ML training or inference.  Proposal 4: For DL beam prediction on AI/ML, the UCI reporting overhead reduction is suggested to be studied.  Proposal 5: Study the potential specification impacts on UCI reporting overhead reduction for DL beam prediction on AI/ML from the following aspects   * Mechanism to facilitate the UCI overhead reduction * New or enhanced signaling/procedure on reporting configuration * Enhanced UCI reporting format including contents, quantization bits number, etc. |
| Intel[13] | Observation 1: The impact of 3GPP specification related procedures for data collection for training as well as inference depends on where the model resides and if training and inferencing is being performed at the same node.  Observation 2: Training dataset construction using 3GPP specified measurement and reporting framework may be advantageous for harmonizing deployment of proprietary AI/ML models. |
| Lenovo[15] | Proposal 7: Study data collection procedure to support both UE-centric and NW-centric AI model training   * For UE-centric model training, study procedure to support UE triggered data collection for model update * For NW-centric model training, support to report larger number of beams in one or more beam report. |
| NEC[16] | Proposal 8: Study the mechanism of obtaining RS specific or dedicated for data collection in model training, model monitoring and model update.  Observation 2: For date collection in model training, model monitoring or model update, the beam information corresponding to input and output (i.e., partial beams in Set A) are needed, rather than the beams information corresponding to all beams in Set A.  Proposal 9: Study the mechanism of beam reporting for data collection in model training, model monitoring and model update. |
| CIACT[17] | Proposal 5: New data type for AI model training needs FFS. |
| CMCC[19] | Proposal 1: For data collection of AI/ML based beam measurement, whether the existing CSI-reporting framework can be reused or any enhancement is needed should be studied. |
| Nokia[20] | Proposal 13: For data collection purpose at the NW side, RAN1 shall further study the CSI reporting enhancement (e.g., reporting more than 4 beams and associated L1-RSRP) such that NW may update the data set for model training/update/fine-tuning. |
| Apple[24] | Proposal 2: If UE position information is used AI/ML aided beam management, user privacy needs to be considered in data collection for model training and input for inference with UE position information. |
| NVIDIA[26] | Proposal 8: For AI/ML model training for beam prediction in spatial/time domain, study potential specification impact related to training data type/size, training data source determination, and assistance signalling and procedure for training data collection. |
| Samsung[27] | Proposal 9. For the data collection for AI/ML model training, in the case that AI/ML model is at gNB-side, the following aspects can be further study:   * Potential enhancement for the measurement and report for data collection * The handling/buffering of the collected data   Proposal 10. For the data collection for AI/ML model training, in the case that AI/ML model is at UE-side, the following aspects can be further study:   * UE report for the preference of data collection, e.g., intended/preferred RS transmission for UE measurement, intended/preferred time domain pattern of the RS transmission * RS measurement configuration for data collection |
| QC[29] | Proposal 2: Study the signalling aspects related to gNB sending assistance information to help UE with data collection for training.  • Examples of such assistance information: information about gNB beam shape, beam boresight directions, 3dB beamwidth, information about gNB antenna array structure, etc. |

###### Proposal 4.3.1

Most of the proposals are made from the high-level perspective. Moreover, the proposals are quite diverging and most of them are mentioned only by one or two companies.

In theory, the procedure and signaling supporting data collection for AI/ML model can be specified in PHY layer and/or higher layer. It is very likely whether L1 or higher layer procedure/signaling is used will be agnostic to the three RAN1 use cases (i.e., CSI, Beam management, positioning). Thus, this issue is recommended to be discussed and determined in Agenda item 9.2.1.

For this agenda item, it seems more efficient to focus on the following aspects in the current stage:

* What should be reported
* What should be configured

The “how” questions can be discussed later.

Therefore, the following proposal is suggested for further discussion.

***Proposal 4.3.1: Regarding the data collection for AI/ML model training at NW side (if the data collection is supported), study the following information for UE reporting as a starting point.***

* ***M L1-RSRPs and the corresponding RS indicator (e.g., CSI-RS, SSB), where M can be larger than 4***
* ***Other information is not precluded***
* ***FFS: the range of M***

|  |  |
| --- | --- |
| Company | Comments |
| Google | We think data collection is used for model monitoring. For model monitoring, L1-RSRP report is not helpful, but UE can directly tell gNB whether the predicted beam can work or not, e.g. the L1-RSRP for the predicted beam is better than current beam or not.  Mod: This proposal is for model training |
| vivo | OK. |
| mediatek | Support |
| NVIDIA | Support |
| Qualcomm | As mentioned in Conclusion 2.2, for the BM use cases the focus is on offline training. Again, for this proposal, the discussions need to wait until 9.2.1 provides a clear understanding of the data collection procedure. If data collection is supposed to be offline, then there is no specification impact involving data collection, so the potential options clearly need to be spelled out in 9.2.1 after which we can come back to this proposal and discuss again.  Mod: By reading the contributions, some companies think even for the offline training, some spec impacts are needed. In order to address the concern, “***(if the data collection is supported)***” is added. We can discuss further the data collection for offline training is needed or not. |
| Apple | Okay |
| Xiaomi | One report for data collection should include beam measurements for more than one time instance. So we suggest the following update  ***Proposal 4.3.1: Regarding the data collection for AI/ML model training at NW side, study the following information for UE reporting as a starting point.***   * ***N time instances’ measurements, where N can be larger than 1*** * ***M L1-RSRPs and the corresponding RS indicator (e.g., CSI-RS, SSB) for each time instance, where M can be larger than 4*** * ***Other information is not precluded*** * ***FFS: the range of M, N*** |
| Samsung | Ok with the proposal in general. |
| LGE | Fine in general, but ***(if the data collection is supported)*** is ambiguous. Even without any additional specification impact, data collection should be done by AI/ML model. |
| NEC | Support |
| CAICT | Fine with FL’s proposal. |
| NTT DOCOMO | Support the proposal. |
| ZTE | We suggest the following generic version:  ***Proposal 4.3.1: Regarding the data collection for AI/ML model training at NW side, study the following information for UE reporting as a starting point.***   * ***FFS: reported quantities, e.g., L1-RSRP, RS indicator (e.g., CSI-RS, SSB)*** * ***FFS: number of reported beams, which can be larger than 4*** |

## AI/ML inference for BM-Case1 & BM-Case2

### General/common aspects

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

|  |
| --- |
| **RAN1#110**  Agreement  In order to facilitate the AI/ML model inference, study the following aspects as a starting point:   * Enhanced or new configurations/UE reporting/UE measurement, e.g., Enhanced or new beam measurement and/or beam reporting * Enhanced or new signaling for measurement configuration/triggering * Signaling of assistance information (if applicable) * Other aspect(s) is not precluded |

The related proposals/observations are copied as below:

|  |  |
| --- | --- |
| FUTUREWEI[1] | *Proposal 4: Study the following aspects as a starting point related to model inference on standards impact.*   * *Enhanced or new configurations/UE reporting/UE measurement, e.g., enhanced or new beam measurement and/or beam reporting* * *Beam indication of the predicted beam(s)* * *Enhanced or new signalling for measurement configuration/triggering* * *Signalling of assistance information (if supported)* * *Other aspect(s) is not precluded*   *Proposal 5: In AI/ML-based beam management, when model training is at NW side and “Alt.3: Beam pair prediction” is adopted, study the standards impact to enable gNB to map the received L1-RSRP measurements to the corresponding Tx-Rx beam pairs.* |
| Huawei[2] | *Similar to our discussions in previous sections, the new UE measurement/reporting can include: larger number of RSRPs reported for Set B as the inference input, or larger number of beam IDs reported as the Top-K of inference output, etc. The enhanced signaling for configuring the AI/ML-based measurement may indicate the relationship between Set A and Set B, e.g., when Set B is a subset of Set A, the mapping relationship between Set B beams/resources to Set A beams/resources. The assistance information should not disclose the proprietary to the other side.* |
| ZTE[3] | *Proposal 11: Study enhanced resource configuration and beam indication mechanism if more flexible triggering or activating approaches are utilized.*  *Proposal 12: Enhanced resource configuration and reporting mechanisms need to be investigated to facilitate the exchange of assistance information, which can be either implicit or explicit.*  *Observation 7: Compared with the beam pair prediction on the network side, the beam pair prediction on the UE side brings less standardization work and does not involve sensitive proprietary information disclosure issues.*  *Proposal 13: To facilitate the AI/ML based Tx-Rx beam pair prediction, enhancements on specification can be studied to support P1 with potentially enhanced beam reporting and indication mechanism.* |
| Spreadtrum[4] | *Observation 3: If AI/ML training is at NW side while AI/ML inference is at UE side, signaling related to AI/ML transfer should be defined.* |
| Vivo[5] | *Observation 13: Assistance information shall be reported to gNB for enhanced beam pair prediction and DL Tx beam prediction with Alt.1.*  *Observation 14: Signaling aspects related to Rx beam or Rx beam pattern indication may be needed in inference stage for Alt.1.*  *Observation 16: Signaling aspects related to the procedure of TCI configuration/indication should be enhanced.*  *Observation 17: Assistance information, such as Tx beam information and/or expected Tx beam information, should be noticed to UE in advance or with resource configuration for option 1 and option 2 with Alt.2 in inference stage.*  *Observation 18: Signaling aspects related to Tx beam or Tx beam pattern request from UE may be needed in inference stage for Alt.2 with considering performance improvement.*  *Observation 20: Signaling aspects related to the procedure of TCI configuration/indication should be enhanced with Alt.2 in inference stage.*  *Observation 21: DL Rx beam prediction in inference stage with Alt.2 has similar specification impacts to training stage, such as P3 resource related information request, and Tx beam information indication.*  *Observation 22: Signaling aspects related to model transfer, model registration/confirmation and model activation should be studied for Alt.3 in inference stage.*  *Proposal 16: For AI/ML model inference, study signaling aspects enhancement related to the procedure of TCI configuration/indication.* |
| IDC[6] | *Observation 8: The current NR specification supporting UE reporting with up to 4 best CRIs/SSBRIs with L1-RSRP or L1-SINR can be very limited for gNB estimation.*  *Proposal 12: Study benefits of simple specification extension of UE reporting.*  *Observation 11: The current NR specification does not consider association between beams with different beam widths.*  *Observation 12: Utilizing association between beams with different beam widths can provide benefits for prediction accuracy e.g., robust estimation/identification of whole spatial characteristics with wide beams and accurate beam identification with narrow beams.*  *Proposal 14: Study benefits of specification enhancements on association between beams with different beam widths.*  *Observation 13: For Rel-15 beam management, actual mapping between DL Tx beam and UE Rx beam is totally based on UE implementation.*  *Observation 14: The implementation-based UE Rx beam selection works for Rel-15, however, UE Rx beam information is crucial to accurately predict beam qualities for AI/ML based beam prediction.*  *Proposal 15: Study benefits of specification enhancements on acquiring UE Rx beam information for DL Tx beam prediction (Alt. 1) and beam pair prediction (Alt. 3).* |
| LGE[9] | *Proposal #1: For the UE AI/ML input, Alt2 can be considered. For the assist information for input, output, training, and inference, consider to express Set A and Set B beams on a pre-defined or configured beam grid.*  *Proposal #2: Consider UE assistance/reporting for determining Set A* |
| Fujitsu[12] | *Observation 1: UCI reporting overhead is increased a lot for DL beam predication on AI/ML training or inference.*  *Proposal 4: For DL beam prediction on AI/ML, the UCI reporting overhead reduction is suggested to be studied.*  *Proposal 5: Study the potential specification impacts on UCI reporting overhead reduction for DL beam prediction on AI/ML from the following aspects*   * *Mechanism to facilitate the UCI overhead reduction* * *New or enhanced signaling/procedure on reporting configuration* * *Enhanced UCI reporting format including contents, quantization bits number, etc.* |
| CMCC[19] | *Proposal 1: The same sort rule of beam pairs is pre-defined so that gNB and UE have the same understanding of index of beam pairs.* |
| NVIDIA[26] | *Proposal 11: For AI/ML based beam prediction in spatial/time domain, study potential specification impact related to report/feedback of model input for inference, type of model input, and model input acquisition and pre-processing.*  *Proposal 12: For AI/ML based beam prediction in spatial/time domain, study potential specification impact related to report/feedback of model inference output and post-processing.* |
| QC[29] | *Proposal 5: For BM-Case1, potential benefits and spec impacts of time-varying set B should be studied*  *Proposal 6: Study the signalling aspects related to beam blockage/failure prediction, as a sub-use case of temporal beam prediction.* |
| Panasonic[30] | *Observation 1: Beam pattern information can be defined as model input to make the model more general. Otherwise, multiple models need to be trained with each corresponding to one specific assumption of Tx beam pattern.*  *Observation 2: For UE-side inference, Tx beam pattern information needs to be made available at UE side.*  *Proposal 4: Study how to efficiently signal the Tx beam pattern information to UE.* |
| KT[31] | *Proposal 3. Study who decides and how to signal Set A and Set B related information for each alternative.* |

###### Proposal 4.4.1.1

Most of the proposals are made from the high-level perspective. Meanwhile, the proposals are quite diverging.

For non-AI based BM, gNB will indicate the beam(s) based on UE reporting. Thus, the indicated beam(s) has been measured by the UE. However, the predicted beam of AI model may be not measured by UE. Thus, a number of contributions suggest to study beam indication for the predicted beam(s) on top of the aspects agreed in the last meeting. The following proposal is suggested for further discussion:

***Proposal 4.4.1.1: In order to facilitate the AI/ML model inference, study the following additional aspect:***

* ***Beam indication of the predicted DL Tx beam(s)***

|  |  |
| --- | --- |
| Company | Comments |
| Google | We failed to see the necessity for this proposal. Beam indication should be the next step after model inference.  Mod: As you said beam indication is needed after model inference, we need to discuss it somewhere. This session seems the best place for discussion. The wording can refined. Let’s wait for more inputs |
| vivo | We are okay to study beam indication. Beam indication is the next step after inference, but it seems inference is the best place to discuss it in the whole LCM.  The other details for inference spec impact have been included in previous sections including input/output, assistant information, etc. |
| mediatek | Support |
| NVIDIA | It would be better to make progress on inference first before discussing beam indication.  Mod: This is for study, not conclude to support it or not. Thus, it should be fine. |
| HW/HiSi | It feels that the proposal is broader than the explanation text before the proposal. The explanation text is for a UE side model that predicts a TX beam which has not been measured, i.e. a beam outside of Set B. If this is the intention of the proposal, we think it should be clarified.  Mod: In my understanding gNB-side model has the similar issue. If I missed something, please feel free to correct me |
| Qualcomm | The proposal needs to be clarified. This does not seem to specifically talk about UE-side or gNB-side AI/ML model, as they have different signaling implications. Also, this is meant to be “predicted DL TX beam(s)”, right?  Mod: Added |
| Apple | Is this for UE-side prediction or NW-side prediction?  Mod: For both. After further discussion, may be the group can find it is not needed for one of them , or not needed for both. |
| Xiaomi | We guess the intention is to talk about the case that if the predicted beam is not a measured beam. In my understanding, it is different between UE-side prediction and NW-side prediction. If it is UE-side prediction, UE can predict the beam that is not measured means that UE knows the predicted beam maybe based on assistance information from gNB. But if it is NW-side prediction, if the predicted beam is not a UE measured beam, UE may not know the predicted beam. The first way to solve this problem is that gNB indicate the relation between measured beam and predicted beam by assistance information to UE. And the second way is that gNB send another RS by the predicted beam and UE performs measurement.  Mod: Understand your comment, but is it related to the proposal? Would you like to elaborate a bit more? In my understanding, the above-mentioned aspects (e.g., the assistance information) are or will be discussed in other proposals/sections. |
| Samsung | Ok with the updated proposal in general. |
| LGE | Fine to study beam indication |
| NEC | Support |
| CAICT | Fine with updated proposal. |
| NTT DOCOMO | As other companies suggested, we prefer to decouple the proposal for UE side and NW side to have a clear picture. |
| ZTE | It depends on whether a second-stage beam sweeping over the predicted top-K beams is conducted. Besides, apart from the differentiation of UE-side model or NW-side model, the beam indication method for spatial domain beam prediction and temporal beam prediction may also be different. |

### AL/ML inference at gNB side

The related proposals/observations for both BM-Case1 and BM-Case2 are copied as below:

|  |  |
| --- | --- |
| FUTUREWEI[1] | *Proposal 5: In AI/ML-based beam management, when model training is at NW side and “Alt.3: Beam pair prediction” is adopted, study the standards impact to enable gNB to map the received L1-RSRP measurements to the corresponding Tx-Rx beam pairs.* |
| ZTE[3] | *Proposal 9: For NW-side beam prediction AI/ML models, enhanced beam reporting mechanisms such as further screening, compression, and reporting of the beam measurement results need to be studied so as to balance the beam prediction performance and beam reporting overhead.* |
| Spreadtrum[4] | *Proposal 4: For beam measurement and reporting, current CSI framework can be considered as starting point.*   * *If AI/ML inference is at NW side, beam reporting needs to be studied to balance the information contained in beam reporting and beam reporting overhead.* * *If AI/ML inference is at UE side, beam reporting needs to be enhanced to report a beam/resource that was not directly measured.*   *Observation 4: For beam indication, the Rel15/16/17 TCI framework can be considered as starting point.*   * *If AI/ML inference is at NW side, how to determine the best Rx beam needs further study* * *If AI/ML inference is at UE side, no specification impact is identified* |
| Vivo[5] | *Observation 15: Report enhancement, including all measured L1-RSRP and/or Rx beam information, is needed for AI model inference at gNB side.*  *Proposal 13: For AI/ML model inference at NW side, at least study the following aspects for potential necessary specification impact:*   * *Signaling aspects enhancement related to Rx beam or Rx beam pattern indication/configuration, for NW side inference* * *Enhance assistance information report related to Rx beam angle, for NW side inference* * *Report enhancement to carry more L1-RSRP and/or Rx beam information, for NW side training* |
| OPPO[7] | *Proposal 11: For BM-Case1 and BM-Case2 when inference at NW side, study beam reporting mechanism on Set B.*  *Proposal 12: For BM-Case1 and BM-Case2 when inference at NW side, study the beam indication mechanism for Tx-Rx beam pair and Rx beam only prediction.* |
| Google[8] | *Proposal 4: When AI/ML model is implemented in the NW side, the output for the AI/ML for spatial domain beam prediction with spec impact should be the reference angle for DL Rx beam refinement (Alt3).*  *Proposal 10: When AI/ML model is implemented in the NW side, the output for the AI/ML for time domain beam prediction with spec impact should be the reference angle for DL Rx beam refinement (Alt3).* |
| Ericsson[10] | *Proposal 9 Consider enhanced UE reporting in line with the evaluation assumptions on set B of beams (e.g. more than 4 beams) to enable NW-sided model inference*  *Proposal 10 Consider mechanism to signal UE assistance data associated with beam measurement report for NW-sided model inference*  *Proposal 11 Consider enhanced UE configurations for NW-sided AI/ML model inference, for example NW indicates potential measurement pre-processing for reducing the UE uplink reporting overhead* |
| CATT[11] | *Proposal 9: Regarding the data collection for BM-Case1 and BM-Case2, if inference is performed at gNB side, the UE needs to report the measurement results (e.g., L1-RSRP) of Set B as model inputs to gNB. Whether beam ID or other assistance information is needed as model inputs should be further studied.*  *Proposal 12: Regarding the model inference for BM-Case1 and BM-Case2, the following aspects should be further studied:*   * *If the model is inferred at gNB side, how to indicate the predicted best beam in TCI states should be studied;* * *If the model is inferred at UE side, how to indicate the N predicted Tx beams to gNB should be studied.* |
| Lenovo[15] | *Proposal 8: Rel-17 CSI reporting framework can be reused for NW-centric beam prediction by increasing the number of beams in a beam report.*  *Proposal 10: Study on how to obtain the assisting information for AI/ML input.* |
| NEC[16] | *Proposal 6: Study the mechanism of reporting more beams in a beam report, e.g., larger than 4.*  *Proposal 7: Study the mechanism of reducing the overhead of beam measurement and reporting in model inference.* |
| Xiaomi[18] | *Proposal 4: For spatial domain beam prediction, study to report Rx beam information, including Rx beam ID/Rx beam shape information of UE to gNB for gNB side inference.*  *Proposal 5: To indicate Rx beam information to UE for obtaining L1-RSRP input to AI/ML model.*  *Proposal 8: Increase the maximum number of beams in beam report for each time instance* |
| CMCC[19] | *Proposal 2: For model inference of BM-Case1 at NW side or UE side, CSI report framework needs further enhancement, including the index of beam pairs and the number of reported beam pairs.* |
| Nokia[20] | *Proposal 14: For model inference at the NW side, RAN1 shall further study the CSI reporting enhancement on how to configure measurements of fixed or variable Set B measurements.* |
| Samsung[27] | *Proposal 2: For BM-Case1, further study the specification impacts for AI/ML inference at NW side considering the following aspects.*   * *Enhancement on L1 beam report mechanism* * *Assistance information for beam prediction*   *Proposal 5: For BM-Case2, further study the specification impacts for AI/ML inference at NW side considering the following aspects.*   * *Enhancement on L1 beam report mechanism* |
| DCM[28] | *Proposal 5: Enhancements on beam measurement reporting should be considered to facilitate or improve the beam prediction at NW side model.*  *Proposal 6: In DL beam prediction with NW-side model, some mechanisms to report Rx beam ID used for beam measurement can be considered as potential specification impacts.*  *Observation 2: Enhancements on beam selection policy in CSI reports might be potential specification impacts for spatial domain beam estimation.*  *Proposal 7: CSI report should be enhanced to improve the performance of time-domain beam prediction, if time-domain beam prediction is supported as sub use-case.* |
| QC[29] | *Proposal 3: For BM-Case1 and BM-Case2, study and evaluate the benefits of beam prediction at UE and gNB and the associated signalling needed to assist or enable beam prediction at each side.*   * + *The trade-off between beam prediction accuracy and required signalling overhead for UE-side and gNB-side inference should be considered in the study.*     - *UE-side inference:*       * *Study enhanced UE L1 report including information from beam prediction*       * *Study signalling aspects related to assistance information from gNB to help beam prediction at UE*     - *gNB-side inference:*       * *Study enhanced UE L1 report to improve beam prediction quality at gNB* |

The related proposals/observations dedicated to BM-Case2 are copied as below:

|  |  |
| --- | --- |
| IDC[6] | *Observation 9: The current NR specification supports measurement restriction to limit UE measurement, however, measurement restriction is to efficiently utilize RS transmissions for multiple beams not to consider time domain characteristics of beam measurement.*  *Observation 10: For gNB which predicts beams by using AI/ML, time domain characteristics of beam measurements are essential as well as spatial domain characteristics.*  *Proposal 13: Study benefits of specification enhancements such as UE reporting with associated time domain information.* |
| LGE[9] | *Proposal #4: For NW-side AI/ML in BM-Case2, consider enhancements on UE reporting and/or beam indication.* |
| NEC[16] | *Proposal 10: Study the mechanism of discontinuous reporting in periodic or semi-persistent beam reporting.*  *Proposal 11: Study the method of indicating the predicted beams and corresponding beam application/dwelling times.* |
| Xiaomi[18] | *Proposal 9: Consider enhancement on beam measurement report to contain more than one time instance.* |

###### Proposal 4.4.2.1

Most of the proposals are made from the high-level perspective. Meanwhile, the proposals are quite diverging. Moreover, many proposals are related to the controversial details of BM-Case1 and BM-Case2, e.g., Rx beam ID, assistance information. For these proposals, it is better to wait for more progress on the details of BM-Case1 and BM-Case2.

Among the dozens of proposals, the enhancement of UE reporting seems supported by a considerable number of companies since the current spec only support to report the measurement results of up to 4 beams. Thus, the following proposal is suggested for further discussion and refinement.

***Proposal 4.4.2.1:******For BM-Case1 and BM-Case2 with a network-side AI/ML model, study the necessity and/or the design of the following UE reporting enhancement for AI/ML model inference***

* ***~~Support~~ UE to report the measurement results of more than 4 beams in one reporting instance***

|  |  |
| --- | --- |
| Company | Comments |
| Google | We are ok with the proposal in principle, but it seems the word “support” in the sub-bullet should be removed since the main-bullet says “study”.  Mod: “support” is removed |
| vivo | OK |
| mediatek | Support |
| NVIDIA | Support |
| HW/HiSi | Support |
| Qualcomm | For gNB-side AI/ML model this is a very narrow example of a possible enhancement, but we believe it should not be limited to this one only at his point. Here’s the suggested proposal:  ***Proposal 4.4.2.1:******For BM-Case1 and BM-Case2 with a network-side AI/ML model, study the necessity and/or the design of the following UE reporting enhancement for AI/ML model inference***   * ***Support UE to report the measurement results of more than 4 beams in one reporting instance*** * ***Consider other L1-report enhancements, e.g., resolution enhancement, compression of L1 report, report of temporal variance of L1-RSRP/L1-SINR measurements, etc.*** |
| Xiaomi | Support and it is better to make it clear that it is enhancement on beam report based on CSI feedback framework by the following update.  ***Proposal 4.4.2.1:******For BM-Case1 and BM-Case2 with a network-side AI/ML model, study the necessity and/or the design of the following ~~UE~~ L1 beam reporting enhancement for AI/ML model inference***   * ***~~Support~~ UE to report the measurement results of more than 4 beams in one reporting instance*** |
| Samsung | Ok with the updated proposal in general. Also fine with the sub-bullet added by QC. |
| LGE | Ok with revision by QC. |
| NEC | Support |
| NTT DOCOMO | Support the proposal. |
| ZTE | For a gNB-side model, the reporting overhead is huge. Thus, enhanced beam reporting mechanisms can be considered, such as further screening and compression of the beam measurement results. Besides, the impact of quantization errors on beam prediction accuracy and how to improve the beam reporting accuracy also need further studied. We suggest the following revision based on QC's update.  ***Proposal 4.4.2.1:******For BM-Case1 and BM-Case2 with a network-side AI/ML model, study the necessity and/or the design of the following UE reporting enhancement for AI/ML model inference***   * ***Support UE to report the measurement results of more than 4 beams in one reporting instance*** * ***Consider other L1-report enhancements, e.g., resolution enhancement, compression of L1 report, reporting format/quantity optimization, report of time-related information of L1-RSRP/L1-SINR measurements, etc.*** |

### AL/ML inference at UE side

The related proposals/observations for BM-Case1 and BM-Case2 are copied as below:

|  |  |
| --- | --- |
| ZTE[3] | *Proposal 10: For UE-side beam prediction AI/ML models, signaling methods need to be investigated to enable recommendation, configuration, and indication of the beam set for measurement.* |
| Spreadtrum[4] | *Proposal 4: For beam measurement and reporting, current CSI framework can be considered as starting point.*   * *If AI/ML inference is at NW side, beam reporting needs to be studied to balance the information contained in beam reporting and beam reporting overhead.* * *If AI/ML inference is at UE side, beam reporting needs to be enhanced to report a beam/resource that was not directly measured.*   *Observation 4: For beam indication, the Rel15/16/17 TCI framework can be considered as starting point.*   * *If AI/ML inference is at NW side, how to determine the best Rx beam needs further study* * *If AI/ML inference is at UE side, no specification impact is identified* |
| Vivo[5] | *Observation 19: Report enhancement, including predicted beam report scheme and/or temporal domain beam report scheme, is needed for AI model inference at UE side.*  *Proposal 14: For AI/ML model inference at UE side, at least study the following aspects for potential necessary specification impact:*   * *Signaling aspects enhancement related to Tx beam or Tx beam pattern request* * *Enhance assistance information configuration/indication related to Tx beam angle and/or expected Tx beam information* * *Report enhancement, including predicted beam report scheme and/or temporal domain beam report scheme*   *Proposal 15: For AI/ML model inference at UE side, further study the following aspects for potential necessary specification impact if P3 is proven beneficial:*   * *P3 training request signaling to gNB* * *P3 resource related information request to gNB, may include Tx beam pattern, minimum number of Tx beam repetitions* |
| OPPO[7] | *Proposal 9: For BM-Case1, if Tx beam or Tx-Rx beam pair is predicted among Set A at UE side, legacy beam reporting and indication mechanism could be reused as a starting point.*  *Observation 3: For BM-Case1, if Rx beam is predicted among Set A at UE side, there seems no strong specification impact.*  *Proposal 10: For BM-Case2, if Tx beam or Tx-Rx beam pair is predicted among Set A at UE side, study how to extend legacy beam reporting and indication for F time instances* |
| Google[8] | *Proposal 5: When AI/ML model is implemented in the UE side, the output for the AI/ML model for spatial domain beam prediction with spec impact should be the reference angle for DL Tx beam refinement (Alt3).*  *Proposal 11: When AI/ML model is implemented in the UE side, the output for the AI/ML model for time domain beam prediction with spec impact should be the reference angle for DL Tx beam refinement (Alt3).* |
| LGE[9] | *Proposal #5: For UE-side AI/ML in BM-Case2, consider enhancements on beam reporting.* |
| Ericsson[10] | *Proposal 12 Study mechanisms of UE reporting in respect to a non-measured beam including a future time instance as a starting point*  *Proposal 13 Study enhanced CSI report configuration to facilitate temporal and spatial beam predictions*  *Proposal 14 The investigation of assistance information signalling should prioritize mechanisms for NW to indicate beam IDs to the UE* |
| CATT[11] | *Proposal 10: Regarding the data collection for BM-Case1 and BM-Case2, if inference is performed at UE side, the UE needs to get the measurement results (e.g., L1-RSRP) of Set B as model inputs. Whether beam ID or other assistance information is needed as model inputs should be further studied.*  *Proposal 12: Regarding the model inference for BM-Case1 and BM-Case2, the following aspects should be further studied:*   * *If the model is inferred at gNB side, how to indicate the predicted best beam in TCI states should be studied;* * *If the model is inferred at UE side, how to indicate the N predicted Tx beams to gNB should be studied.* |
| Intel[13] | *The specification impact may be from UE triggered CSI-RS transmissions for beam measurements.* |
| Lenovo[15] | *Proposal 9: Rel-17 CSI reporting framework can be reused for UE-centric beam prediction by configuring measurement beam Set B as the channel measurement resource but the reported beam is selected from another prediction beam Set A.* |
| Xiaomi[18] | *Proposal 3: For spatial domain beam prediction, study how to indicate the Tx beam information, including Tx beam ID/Tx beam shape information of gNB to UE for UE side inference.* |
| CMCC[19] | *Proposal 2: For model inference of BM-Case1 at NW side or UE side, CSI report framework needs further enhancement, including the index of beam pairs and the number of reported beam pairs.* |
| Nokia[20] | *Observation 21: For DL Tx beam or DL Tx-Rx beam pair prediction at the UE with collaboration level-y,*   * *the exchanged collaboration signaling may be required to carry assistant info related to NW’s beam pattern layout. Such assistant info may be used for model input.* * *the exchanged collaboration signaling may be required to consider the details such as selecting a model at the UE or indicating details related to model management.*   *Proposal 15: For UE side DL Tx beam or DL Tx-Rx beam pair prediction with collaboration level-y, RAN1 shall investigate further details about UE side model generalization and the corresponding NW-UE model alignment scheme.*  *Proposal 16: For UE side DL Tx beam or Tx-Rx beam pair prediction, further study configuring different RS resource sets for beam prediction and beam measurements.*  *Proposal 17: For UE side DL Tx beam or Tx-Rx beam pair prediction, further required changes on CSI reporting quantities to report predicted beams.*  *Observation 22: For BM-Case2, UE side DL Tx beam or Tx-Rx beam pair prediction, the “top-N” beams CSI measurement/report configuration and the TCI configuration may have potential spec impact.* |
| Samsung[27] | *Proposal 3: For BM-Case1, further study the specification impacts for AI/ML inference at UE side considering the following aspects.*   * *Assistance information for AI/ML inference at UE side* * *Enhancement on L1 beam report mechanism*   *Proposal 6: For BM-Case2, further study the specification impacts for AI/ML inference at UE side considering the following aspects.*   * *Enhancement on L1 beam report mechanism* * *UE-side case/events that can leverage the predicted/future L1-RSRP* |
| DCM[28] | *Proposal 8: Mechanisms to provide DL Tx beam information from NW to UE could be potential specification impacts in DL beam prediction*  *Observation 4: Boresight direction and/or (relative) power per angle for each reference signal can be potential assistance information of Tx beam in DL beam prediction.*  *Observation 5: Some enhancements of beam measurement reporting can be potential specification impacts for temporal DL beam prediction with UE side model.* |
| QC[29] | *Proposal 3: For BM-Case1 and BM-Case2, study and evaluate the benefits of beam prediction at UE and gNB and the associated signalling needed to assist or enable beam prediction at each side.*   * + *The trade-off between beam prediction accuracy and required signalling overhead for UE-side and gNB-side inference should be considered in the study.*     - *UE-side inference:*       * *Study enhanced UE L1 report including information from beam prediction*       * *Study signalling aspects related to assistance information from gNB to help beam prediction at UE*     - *gNB-side inference:*       * *Study enhanced UE L1 report to improve beam prediction quality at gNB*   *Proposal 4: Regarding the sub use case BM-Case1 and BM-Case2, study the following potential signalling enhancements for UE-side inference:*   * *L1-report enhancement to report Tx beam ID(s) and/or the predicted L1-RSRP(s) of the N predicted DL Tx beams and/or other information (e.g., probability for the beam to be the best beam, the associated confidence, beam application time/dwelling time, predicted beam failure, etc.)*   + *Note: the enhanced L1-report may distinguish between measured and predicted L1-RSRP(s)*   + *The predicted L1-RSRP(s) may be for a single or multiple future prediction instances* * *L1-report enhancement to report Tx beam angle(s) and/or the predicted L1-RSRP(s) of the N predicted DL Tx beams*   + *E.g., N predicted beams can be the top-N predicted beams*   + *FFS: details of Tx beam angle(s), e.g., channel AoA/AoD and/or other parameters*   *Proposal 6: Study the signalling aspects related to beam blockage/failure prediction, as a sub-use case of temporal beam prediction.* |

###### Proposal 4.4.3.1

Most of the proposals are made from the high-level perspective. Meanwhile the proposals are quite diverging and most of the proposals are mentioned by only one or two companies.

Among the dozens of proposals, the enhancement of UE reporting to support un-measured beam seems supported by a considerable number of companies since the current spec only supports to report the measured beam(s). Meanwhile, if Set B is a sub set of Set A, then some of the predicted beam(s) may be the measured beam(s). Thus, the following proposal is suggested for further discussion and refinement.

***Proposal 4.4.3.1:******For BM-Case1 and BM-Case2 with a UE-side AI/ML model, study the necessity and/or design of L1 signaling to report the following information of AI/ML model inference***

* ***The beam(s) that is based on the output of AI/ML model inference~~, and that may be not measured by UE~~***
* ***Predicted L1-RSRP corresponding to the beam(s)***
* ***FFS: other information***

|  |  |
| --- | --- |
| Company | Comments |
| Google | We do not think UE needs to report L1-RSRP for the predicted beam as UE may not have measured this beam.  Mod: Let’s wait for more inputs |
| vivo | We are not sure why the report needs to include the indication of beams not measured by UE. gNB configures the beams to be measured by UE and UE reports some beams based on the output. gNB should know which reported beam are not included in set B.  Further, the reported beams should be beams based on output of the AI/ML model. It may not just be the exact output of the AI/ML model.  Mod: 1. “***, and that may be not measured by UE***” is removed to avoid the confusion  2. “based on” is added |
| mediatek | Support |
| NVIDIA | Support |
| qualcomm | For UE-side inference, on top of the ‘enhanced L1 report’ the signaling of assistance information from gNB to UE to help UE with prediction should also be considered.  Mod: If understand correctly, it has been covered with the previous agreement “Signaling of assistance information (if applicable)” except the agreement doesn’t say it is L1 signaling |
| Xiaomi | For BM case2, it is possible to predict the beam of more than one future time instance. So we suggest the following update  ***Proposal 4.4.3.1:******For BM-Case1 and BM-Case2 with a UE-side AI/ML model, study the necessity and/or design of L1 signaling to report the following information of AI/ML model inference***   * ***The beam(s) of N future time instance that is based on the output of AI/ML model inference, where N can be equal to or larger than 1.~~, and that may be not measured by UE,~~*** * ***Predicted L1-RSRP corresponding to the beam(s)*** * ***FFS: other information*** |
| Samsung | Ok with the proposal in general. |
| LGE | Support |
| NEC | Support |
| CAICT | Support |
| NTT DOCOMO | Support the proposal. |
| ZTE | We agree with xiaomi that time-related information may need to be reported if beams of multiple future time instances are predicted. |

## Model monitoring

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

|  |
| --- |
| **RAN1#110**  Agreement  Regarding the model monitoring for BM-Case1 and BM-Case2, to investigate specification impacts from the following aspects   * Performance metric(s) * Benchmark/reference for the performance comparison * Signaling/configuration/measurement/report for model monitoring, e.g., signaling aspects related to assistance information (if supported), Reference signals * Other aspect(s) is not precluded |

The related proposals/ observations are copied as below:

|  |  |
| --- | --- |
| Huawei[2] | *Proposal 11: For the spec impact of model monitoring, RAN1 studies the following options for performance metrics:*   * *Intermediate results, e.g., predicted beam/RSRP accuracy.* * *Eventual KPI, e.g., RSRP, throughput, etc.*   *Proposal 12: For the spec impact of model monitoring, consider the following operation modes for monitoring:*   * *NW monitoring mode, where UE reports the measurement results (e.g., RSRPs, predicted beam ID, best beam ID) to NW, and NW makes the monitoring decisions (e.g., model activation/deactivation/updating/switching).* * *Joint monitoring of NW and UE, where UE performs measurement, calculates performance metrics (e.g., predicted beam/RSRP accuracy) and reports to NW, and NW makes monitoring decisions.* * *UE monitoring mode, where UE performs measurement, calculates performance metrics and makes monitoring decisions* |
| ZTE[5] | *Observation 8: Model monitoring can be performed by comparing the predicted optimal beam with the realistic optimal beam, which is obtained by measuring one or more resource sets consisting of CSI-RS or alternatively SS blocks that correspond to different downlink beams at the whole beam space.* |
| Spreadtrum[4] | *Observation 5: Considering the reference for the performance comparison,*   * *If set A is used as the reference, UE reporting overhead may be significant.*   *- If set B is used as the reference, only part of the output results will be compared.*  *Proposal5: Actual RSRP of Set A/ Set B used as the reference for the performance comparison needs to be further studied.*  *Proposal6: The RSRP difference evaluated by comparing actual RSRP and predicted RSRP can be used as a performance metric.* |
| Vivo[5] | *Proposal 18: Study specification impact of model performance monitoring for both spatial domain and temporal domain beam prediction regarding at the following aspects:*   * 1. *Monitoring configuration and/or activation conditions*   2. *Monitoring resources, e.g. reference beam pattern*   3. *Monitoring metrics, e.g. beam prediction accuracy, top-k predicted RSRP difference*   4. *Beam report enhancement with monitored results reporting*   5. *Impairments for monitoring, e.g., how to monitor with non-ideal labels* |
| OPPO[7] | *Proposal 13: Study the performance monitoring mechanism of AI/ML model for beam prediction.* |
| Google[8] | *Proposal 6: For spatial domain beam prediction, the beam quality for current beam from an indicated TCI can be used for performance validation, and if none of the predicted beam(s) can provide better beam quality than current beam, the predicted beam(s) are assumed to fall to pass the performance validation.*  *Proposal 12: For time-domain beam prediction, the beam quality for current beam from an indicated TCI can be used for performance validation, and if none of the predicted beam(s) can provide better beam quality than current beam, the predicted beam(s) are assumed to fall to pass the performance validation.*  *Proposal 13: Study UE feedback before the beam action time for performance validation for predicted beam in addition to the ACK/NACK for the TCI update signaling.* |
| Ericsson[10] | *Observation 2 No specification impact foreseen for NW-sided model monitoring*  *Proposal 15 Study performance metrics comprising both per-sample prediction error and statistical metrics*  *Proposal 16 Study the following potential standard impact for monitoring the UE-sided model performance for the considered beam management use cases:*   * 1. *triggering conditions for model monitoring*   2. *mechanisms to support UE reporting its model performance related metric to the NW*   3. *mechanisms to support NW indicating the model monitoring results to the UE.* |
| CATT[11] | *Proposal 13: Regarding the model monitoring for BM-Case1 and BM-Case2, the performance metric(s) can be the beam prediction accuracy related KPIs, e.g., beam prediction accuracy (%) for Top-1 and/or Top-K beams.*  *Proposal 14: Regarding the model monitoring for BM-Case1 and BM-Case2, the benchmark/reference for the performance comparison can be the R15/R16 legacy mechanism using to calculate the best beam in Set A.*  *Proposal 15: Regarding the model monitoring for BM-Case1 and BM-Case2, which side takes responsibility on model monitoring, e.g. UE side or gNB side, should be studied.*  *Proposal 16: Regarding the model monitoring for BM-Case1 and BM-Case2, the spec impacts of following procedures based on model monitoring results should be studied, e.g., model update/switching/fallback.* |
| Intel[13] | *Proposal 2: Model monitoring should be defined specifically per use-case and its performance and impact may vary across different use-cases* |
| NEC[16] | *Proposal 5: Study the direct or indirect metrics and mechanisms of model monitoring.* |
| Xiaomi[18] | *Proposal 10: gNB to transmit all beams in set A periodically/semi-persistently/ a-periodically for performance monitoring.*  *Proposal 11: Threshold of beam prediction accuracy related KPIs can be used for performance monitoring.* |
| CMCC[19] | *Proposal 3: For model inference of BM-Case1, beam prediction accuracy related KPI can be used as the metric of model performance monitoring.*  *Proposal 4: For model monitoring of BM-Case 1, the procedure of NW-based model monitoring can be studied for both NW-sided and UE-sided model.* |
| Nokia[20] | *Observation 23: The wrong beam prediction of the network can cause beam failures or radio link failure which can increase the service interruption time and signaling overhead to handle the recovery of the connection.*  *Proposal 18: For the NW-sided beam prediction, further study the model monitoring by considering frequent measurement and reporting of Set A, and using it to derive model performance metrics.*  *Observation 24: UE can be configured to compare the predicted beam IDs (or predicted beam RSRPs) to the actual beam measurements from Set A with certain rules (i.e. RSRP prediction error or prediction accuracy under/higher certain threshold) to monitor DL TX beam or DL Tx-Rx beam pair prediction failure.*  *Proposal 19: For the UE-sided beam prediction, further study the model monitoring by considering frequent measurement and reporting of Set A, and associated specification impacts.* |
| NVIDIA[26] | *Proposal 10: For AI/ML based beam prediction in spatial/time domain, study potential specification impact related to assistance signalling and procedure for model performance monitoring and model update/tuning.* |
| Samsung[27] | *Proposal 11. For the performance metric(s) of AI/ML model monitoring, beam prediction accuracy related KPIs agreed in 9.2.3.1 can be considered.*  *Proposal 12. For benchmark/reference for the performance comparison of the AI/ML model monitoring, baseline performance options for spatial-domain beam prediction and temporal beam prediction agreed in 9.2.3.1 can be considered.*  *Proposal 13. For the AI/ML model monitoring, in the case that AI/ML model is at gNB-side, the following aspect can be further study:*   * *Potential enhancement for the measurement and report for model monitoring*   *Proposal 14. For the AI/ML model monitoring, in the case that AI/ML model is at UE-side, the following aspects can be further study:*   * *UE report for the performance/validation of AI/ML model* |
| DCM[28] | *Proposal 3: The following real time performance metrics should be considered for DL beam prediction.*  *・Model accuracy (e.g. L1-RSRP difference, beam accuracy probability)*  *・Upper bound of beam prediction gain*  *・Expected beam prediction gain compared to non-beam prediction*  *Observation 1: The following values are necessary for calculating model accuracy, upper bound of beam prediction gain, and beam prediction gain compared to non-beam prediction.*  *・Model inference results: predicted beam quality of Set A (e.g. estimated L1-RSRP of Set A)*  *・Ground truth data: actual beam quality of Set A (e.g. L1-RSRP of Set A)*  *・Inputs of model inference: beam measurements of Set B (e.g. L1-RSRP of Set B)*  *Proposal 4: Consider RS configuration to enable both Set A and Set B beam measurement with the following condition.*  *・Spatial domain beam prediction: SetA and SetB beam measurements at close time*  *・Temporal beam prediction: SetA and SetB beam measurements with certain prediction time offset*  *Observation 3: Performance metric calculation at UE side requires new signalling mechanism to receive model inference results from NW, while performance metric can be calculated at NW side with the existing beam measurement reports*  *Observation 6: it is more straightforward to calculate some performance metrics at UE side than NW side for the DL beam prediction with UE side model.* |
| QC[29] | *Proposal 7: For BM-Case1 and BM-Case2, study the signalling aspects related to exchanging information about beam prediction quality and a metric for beam prediction quality*   * *Study the impact of beam prediction quality on activating/deactivating AI/ML module at UE.*   *Proposal 8: For BM-Case1 and BM-Case2, study the signalling aspects related to gNB sending assistance signalling to help UE in comparing predicted measurements with actual measurements.*   * *This assistance signalling can be in the form of auxiliary reference signals.* |

Many companies think that the monitoring procedure will be different when the AI/ML model inference is performed in different side (i.e., UE side, or NW side). Thus, the following discussions will be separate for NW-side model and UE-side model, respectively.

### NW-side model

###### Proposal 4.5.1.1

For a network-side AI/ML model, it is nature to support the model monitoring at network side. Thus, the following proposal is suggested for further discussion and refinement.

***Proposal 4.5.1.1:******For BM-Case1 and BM-Case2 with a network-side AI/ML model, support the model monitoring performed at network side.***

|  |  |
| --- | --- |
| Company | Comments |
| Google | We think the model monitoring should contain two steps: 1) predicted beam quality monitoring; 2) determining whether current model still works. The first step is better to be in UE side, since the UE knows the beam quality. The second stop can be in either NW side or UE side like beam indication and beam failure recovery.  Therefore, model monitoring should not be a single-side operation.  Mod: understand the intention. Let’s wait for more inputs and then see how to modify the proposal |
| vivo | For specification impact of model monitoring, shouldn’t we focus on UE monitoring as NW monitoring should be implementation? UE monitoring is the motivation to have new signaling (e.g., reporting) or new UE measurement behavior.  Mod: Companies have different views on whether NW monitoring is implementation or not. We can study it and then make some conclusion whether it is implementation or not. That’s is why Proposal 4.5.1.2 is suggested to collect companies’ view. |
| mediatek | Support, but UE-based monitoring can also be studied. |
| NVIDIA | Support |
| HW/HiSi | Support |
| Xiaomi | We are not clear about the definition of network-side monitoring. In my understanding for model monitoring, even it is a NW-side model, UE need to perform measurement on beams in set B and set A, and report the measurements of set B for model input together with the measurements of top-K beams based on the measurements of set A for performance monitoring. Thus there is some new UE behavior. If the definition of network-side monitoring is that network determines the model is valid or not based on the measurement report from the UE, we are OK for this proposal.  Mod: In my understanding, the comments are talking about similar thing as google. Let’s wait for more inputs and try to find better way to discuss this issue |
| Samsung | In our understanding, it is a little bit rush to support this feature directly. Some wording suggestion is as follows.  ***Proposal 4.5.1.1:******For BM-Case1 and BM-Case2 with a network-side AI/ML model, support to study the model monitoring performed at network side.*** |
| LGE | If this is about basic assumption, which may not have any specification impact in the end, the proposal is fine to us. |
| NEC | We would like to clarify that, model monitoring at NW side means that making decision (e.g., model switching/updating/deactivation) based on the monitoring results at NW side, right? |
| CAICT | Support |
| NTT DOCOMO | As Google mentioned, the model monitoring process is not so clear. To clarify, we prefer updating the proposal into two proposals as follow  ***Proposal 4.5.1.1-a:******For BM-Case1 and BM-Case2 with a network-side AI/ML model, support the performance metric calculation at NW side in model monitoring.***  ***Proposal 4.5.1.1-b:******For BM-Case1 and BM-Case2 with a network-side AI/ML model, support the performance comparison (e.g., determine whether the model provides better performance than fallback approaches or determine whether current model still works) at NW side based on the calculated performance metric in model monitoring.*** |
| ZTE | Support the clarification from Google and DCM. For a network-side AI/ML model, at least the finally decision on whether current model still works should be made at NW side. |

###### Proposal 4.5.1.2

Some companies think there is no spec impact if the model monitoring is performed at network side for a NW-side model. In contrast, some other companies think there may be some spec impact dedicated to the model monitoring. Thus, the following proposal is to collect the potential spec impacts for further discussion and down-selection.

***Proposal 4.5.1.2:******When the model monitoring is performed at network side for a network-side AI/ML model, study the necessity and/or specification impacts from the following aspects as a starting point***

* **…**

|  |  |
| --- | --- |
| Company | Comments |
| Google | As commented above, this model monitoring should not be a single-side operation. |
| Xiaomi | Share same view as google that new UE behavior is needed. |
| Samsung | We think the enhancement is still needed when the model monitoring is performed at network side for a network-side AI/ML model. For example, the beam measurement and report of UE can be enhanced to efficiently acquire the results of set A for model monitoring.  ***Proposal 4.5.1.2:******When the model monitoring is performed at network side for a network-side AI/ML model, study the necessity and/or specification impacts from the following aspects as a starting point***   * ***Beam measurement and report for model monitoring*** |
| NTT DOCOMO | Share the similar view with Samsung. Some enhancements can be considered even in model monitoring at NW side, e.g., beam measurement reporting for NW side. |
|  |  |
|  |  |

### UE-side model

###### Collect views 4.5.2.1

For a UE-side AI/ML model, there are different views on at which side (i.e., UE side, or network side) the model monitoring should be performed:

* Some companies think it is natural for UE to do the monitoring.
* Some other companies believe that network should be responsible to the model monitoring.

Companies are invited to provide views on the different alternatives.

|  |  |  |
| --- | --- | --- |
| Model monitoring for UE-side model | | |
|  | Support | Not support |
| Atl1. Model monitoring is performed at UE side |  |  |
| Atl2. Model monitoring is performed at network side |  |  |
| Both Alt.1 and Alt.2 |  |  |

Companies can provide detailed inputs (if any) in the following table.

|  |  |
| --- | --- |
| Company | Comments |
| Google | As we commented before, model monitoring should contain two steps: 1) predicted beam quality monitoring; 2) determining whether current model still works. The first step is better to be in UE side, since the UE knows the beam quality. The second stop can be in either NW side or UE side like beam indication and beam failure recovery. |
| vivo | We support to focus on UE side monitoring as commented above. |
| mediatek | UE-side model monitoring can also be supported. |
| NVIDIA | Support to leave options open for now and continue the study. |
| Xiaomi | Support to study both options first. |
| Samsung | In our view, the decision related to model monitoring should be controlled by gNB. For example, gNB indicates the model update/fallback. In the meanwhile, UE can provide the information related to model monitoring. |
| CAICT | We would like to consider both Alt.1 and Alt.2. |
| NTT DOCOMO | We prefer to subdivide the processes of model monitoring as a first step. |
| ZTE | Open to discuss. The gNB can judge whether the performance of the deployed UE-side model is reliable according to UE reporting contents. Or, more directly, the UE notifies the gNB of the model monitoring result implicitly or explicitly in a reporting instance. Based on the model monitoring result, the gNB performs further operations such as AI model deactivation, model switching, and fallback. |

## Capability

The related proposals/ observations are copied as below:

|  |  |
| --- | --- |
| FUTUREWEI[1] | *Proposal 7: Regarding AI/ML-based beam management, study the standards impact, including AI/ML related UE configuration/capability reporting, related to AI/ML model selection/configuration (like activation/deactivation) in case multiple trained AI/ML models are deployed.* |
| Huawei[2] | *Proposal 13: Study the potential specification impact for UE capability, including the following aspects as a starting point: data collection, model training, inference latency, monitoring, models switching, model updating. Details can be discussed until further progress has been made for schemes themselves and their related spec impact.* |
| OPPO[7] | *Proposal 14: For both BM-Case1 and BM-Case2, study which content(s) should be included as UE capability for beam prediction.* |
| Lenovo[15] | *Proposal 6: Study UE/NW capability related signaling corresponding to AI-based beam management under different network-UE collaboration levels.* |
| NVIDIA[26] | *Proposal 13: For AI/ML based beam prediction in spatial/time domain, study potential specification impact related to UE capability for AI/ML based beam prediction including model training, model inference and model monitoring.* |

**Mod recommendation**: TBD

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

# Summary of Discussion

## Proposal for 1st GTW

### Conclusion 3.5.2a

***Conclusion 3.5.2a:*** ***For the determination/selection of assistance information (if supported),***

* ***The proprietary/privacy information should not be disclosed***
* ***The performance and specification impact should be considered***
  + ***~~Assistance information can be considered when there is some corresponding evaluation result(s).~~***

To LGE/Huawei: the last sub-bullet is removed.

To vivo: The generalization is included in the performance. For example, the EVM FL summary has the following description on generalization.

|  |
| --- |
| 3. 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 |

### Conclusion 3.7a

* BM-Case3 got 6 supporting companies (MTK, Google, Sony, Apple, IDC, Fujitsu). Other sub use cases get fewer supporting companies.
* 15 companies (LGE, ZTE, NEC, CAICT, NVIDIA, FUTUREWEI, Xiaomi, Spread, vivo, QC, Fujitsu, HW, DCM, SS, CMCC) suggested to deprioritize all other sub use cases.

***Conclusion 3.7a:***

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

To google: In Mod’s understanding, BM case3/4/6/9 are not included in BM-Case1 since we only agreed to support Case1 / Case2 and these cases are listed separately.

To CATT, Sony: In Mod’s understanding, BM case3 are not included in BM-Case1 since we only agreed to support Case1 / Case2 and this case is listed separately.

To Nokia: the highlighted part captured the suggestion

### Proposal 2.1a

* Alt.1 is supported by 26 companies
* Alt.2 is supported by 25 companies, not supported by 1 company
* Alt.3 is supported by 7 companies, not supported by 16 companies
* Alt.4 is supported by 2 companies, not supported by 23 companies.

***Proposal 2.1a: For the sub use case BM-Case1 and BM-Case2, at least support Alt.1 and Alt.2 for AI/ML model training and inference for further study:***

* ***Alt.1. AI/ML model training and inference at NW side***
* ***Alt.2. AI/ML model training and inference at UE side***
* ***FFS: Alt.3. AI/ML model training at NW side, AI/ML model inference at UE side***
* ***~~Regarding whether to support Alt.3 for BM-Case1 and BM-Case2, wait for the conclusion/agreement of Agenda item 2.9.1 on whether to support mode transfer or not~~***
  + ***~~Alt.3. AI/ML model training at NW side, AI/ML model inference at UE side~~***

To Huawei: In moderator’s understanding, since only one-sided model was agreed for further study so far, the follow-up discussions are only for one-sided model unless the group agree to support the study of two-sided model.

# Reference

1. R1-2208369 Continued discussion on other aspects of AI/ML for beam management FUTUREWEI
2. R1-2208432 Discussion on AI/ML for beam management Huawei, HiSilicon
3. R1-2208524 Discussion on other aspects for AI beam management ZTE
4. R1-2208550 Discussion on other aspects on AIML for beam management Spreadtrum Communications
5. R1-2208637 Other aspects on AI/ML for beam management vivo
6. R1-2208683 Discussion for other aspects on AI/ML for beam management InterDigital, Inc.
7. R1-2208853 Other aspects of AI/ML for beam management OPPO
8. R1-2208881 On Enhancement of AI/ML based Beam Management Google
9. R1-2208902 Other aspects on AI/ML for beam management LG Electronics
10. R1-2208907 Discussion on AI/ML for beam management Ericsson
11. R1-2208970 Discussion on AI/ML for beam management CATT
12. R1-2209014 Sub use cases and specification impact on AI/ML for beam management Fujitsu
13. R1-2209050 Use-cases and Specification Impact for AI/ML beam management Intel Corporation
14. R1-2209096 Consideration on AI/ML for beam management Sony
15. R1-2209123 Further aspects of AI/ML for beam management Lenovo
16. R1-2209146 Discussion on AI/ML for beam management NEC
17. R1-2209233 Discussions on AI-ML for Beam management CAICT
18. R1-2209280 Discussion on other aspects on AI/ML for beam management xiaomi
19. R1-2209331 Discussion on other aspects on AI/ML for beam management CMCC
20. R1-2209370 Other aspects on ML for beam management Nokia, Nokia Shanghai Bell
21. R1-2209391 Discussions on Sub-Use Cases in AI/ML for Beam Management TCL Communication
22. R1-2209402 Discussion on other aspects on AI/ML for beam management ETRI
23. R1-2209509 Other aspects on AI/ML for beam management MediaTek Inc.
24. R1-2209579 Other aspects on AI/ML for beam management Apple
25. R1-2209614 Discussion on AI/ML for beam management Rakuten Symphony
26. R1-2209628 AI and ML for beam management NVIDIA
27. R1-2209725 Representative sub use cases for beam management Samsung
28. R1-2209899 Discussion on AI/ML for beam management NTT DOCOMO, INC.
29. R1-2209979 Other aspects on AI/ML for beam management Qualcomm Incorporated
30. R1-2210085 Discussion on sub use cases of AI/ML beam management Panasonic
31. R1-2210086 Discussion on other aspects on AI/ML for beam management KT Corp.
32. R1-2209978 Evaluation on AI/ML for beam management Qualcomm Incorporated

# Appendix A: Contact Information

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

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

# Appendix B: Agreements

## RAN1#110bis-e

## RAN1#110

Agreement

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

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

Agreement

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

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

Agreement

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

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

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

Agreement

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

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

Agreement

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

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

Agreement

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

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

Agreement

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

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

Agreement

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

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

## RAN1#109-e

Agreement

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

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

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

Agreement

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

* The value of K is up to companies

Agreement

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

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

Agreement

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

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

Agreement

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

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

Conclusion

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

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

Conclusion

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

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

Conclusion

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

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

Conclusion

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

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