**3GPP TSG RAN WG1 Meeting #110 R1-220xxxx**

**Toulouse, France, August 22 – 26, 2022**

**Source:** Moderator (Samsung)

**Title:** Feature lead summary #0 evaluation of AI/ML for beam management

**Agenda Item:**  9.2.3.1

**Document for:** Discussion and Decision

# Introduction

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

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

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

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

In this contribution summarized the discussions and proposal on evaluation methodology (EVM) and KPIs from contributions submitted to AI 9.2.3.1 for beam management (BM).

The issues in this document are tagged and color coded with High Priority or Medium Priority. The issues that are in the focus of this round of the discussion are furthermore tagged FL1.

Follow the naming convention in this example:

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

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

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

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

#### FL1 Question 0-1

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

|  |  |  |
| --- | --- | --- |
| Company | Point of contact | Email address |
| OPPO | Jianfei (Jeffrey) Cao | caojianfei@oppo.com |
| MediaTek | Gyu Bum Kyung | gyubum.kyung@mediatek.com |
| CAICT | Xiaofeng Liu | Liuxiaofeng1@caict.ac.cn |
| Xiaomi | Mingju LI | limingju@xiaomi.com |
| vivo | Hao Wu | hao.wu@vivo.com |
| FUTUREWEI | Baoling Sheen | bsheen@futurewei.com |
| CATT | Da Wang | wangda@catt.cn |
| CMCC | Jun Zuo | zuojun@chinamobile.com |
| ZTE, Sanechips | Wenfeng Liu | liu.wenfeng@zte.com.cn |
| NTT DOCOMO | Haruhi Echigo  Liu Liu | [haruhi.echigo.fw@nttdocomo.com](mailto:haruhi.echigo.fw@nttdocomo.com)  liul@docomolabs-beijing.com.cn |
| Ericsson | Henrik Ryden | henrik.a.ryden@ericsson.com |
| Samsung | Jackson Wang  Yeon-Geun Lim | [h0809.wang@samsung.com](mailto:h0809.wang@samsung.com)  yg.lim@samsung.com |
| Lenovo | Venkata Srinivas Kothapalli | vkothapalli@lenovo.com |
| Qualcomm | Hamed Pezeshki | hamedp@qti.qualcomm.com |
| HW/HiSi | Thorsten Schier | Thorsten.schier@huawei.com |

# Evaluation methodology on AI/ML in beam management

## 1.1 Evaluation assumptions

### Open issues on evaluation assumption of SLS

The following table summarized the companies’ proposals/views on the open issues for evaluation assumption of SLS:

|  |  |  |
| --- | --- | --- |
| **Parameters** | **Values** | **Companies’ views** |
| **UE Speed** | * For spatial domain beam prediction, 3km/h * For time domain beam prediction: 30km/h (baseline), 60km/h (optional) * Other values are not precluded | * Huawei [2]: Proposal 6: For AI/ML-based temporal domain beam prediction evaluation, higher UE speed value(s), e.g., 90km/h or 120km/h, should be considered in addition to 30km/h. * DoCoMo [26]: Proposal 4: UE speed 60km/h or higher should be considered as the baseline assumption for performance evaluation. |
| **UE distribution** | * FFS UEs per sector/cell for evaluation. More UEs per sector/cell for data generation is not precluded. | * 10 UEs per sector/cell * Vivo[3], ZTE[4], Nokia[21], Interdigital [6](20 UEs optional), Intel [13](for UTP), Ericsson [20](UTP) * 1 UE per sector/cell * Samsung [17] (for UTP, more for AI training/testing) * 10% as the number of training UEs * Google [7]   Others:   * Ericsson [20]: * Number of UEs per/sector per cell during data collection presented if relevant |
| * For spatial domain beam prediction: FFS: * Option 1: 80% indoor ,20% outdoor as in TR 38.901 * Option 2: 100% outdoor * For time domain prediction: 100% outdoor | * Option 1: 80% indoor ,20% outdoor as in TR 38.901 * ZTE [4], Interdigital [6], CATT [10], Intel [13], xiaomi[14] * Option 2: 100% outdoor * Vivo [3], Intel [13], Samsung [16], Nokia [20], DoCoMo [26] |
| **BS Antenna Configuration** | [One panel: (M, N, P, Mg, Ng) = (4, 8, 2, 1, 1), (dV, dH) = (0.5, 0.5) λ as baseline]  [Four panels: (M, N, P, Mg, Ng) = (4, 8, 2, 2, 2), (dV, dH) = (0.5, 0.5) λ. (dg,V, dg,H) (2.0, 4.0) λ as optional]  Other assumptions are not precluded.    Companies to explain TXRU weights mapping.  Companies to explain beam selection.  Companies to explain number of BS beams | * (M, N, P, Mg, Ng) = (4, 8, 2, 1, 1), (dV, dH) = (0.5, 0.5) λ * ZTE [4], Google [7], CATT [10], Xiaomi [14], Samsung [17], Nokia [21] * One panel: (M, N, P, Mg, Ng) = (8, 8, 2, 1, 1) * Intel [13] |
| **UE Antenna Configuration** | * [Panel structure: (M,N,P) = (1,4,2)] * 2 panels (left, right) with (Mg, Ng) = (1, 2) as baseline * Other assumptions are not precluded     Companies to explain TXRU weights mapping.  Companies to explain beam and panel selection.  Companies to explain number of UE beams | * Panel structure: (M,N,P) = (1,4,2) * ZTE[4], Google[7], CATT[10], Samsung[17] * [Mg Ng M N P] = [1 1 2 4 2], [dV, dH] = [0.5,0.5] λ * China Telecom[16] |
| **Traffic Model** | FFS:   * Option 1: Full buffer * Option 2: FTP model   Other options are not precluded | * Interdigital [6]: For traffic model, support the following evaluation assumptions: * For beam information related KPIs, no traffic model is needed to be defined as UE is only measuring reference signals not decoding actual PDSCHs. * For system performance related KPIs, FTP traffic should be used to reflect practical traffics for the evaluation. * Samsung [17]: * Option 1: Full buffer * Other options are not precluded * Ericsson [20] * Agree to use full buffer and FTP-model(s) for system level evaluations as a starting point. FTP model parameters are FFS |
| **BS Tx Power** | [40 dBm] | CATT [10], Samsung [17] |

Based on the above views, the following proposals can be discussed:

#### FL1 (High) Question 1-1-1a

**Proposal 1-1-1a: (**updated based on the agreements in RAN 1 #109-e**):**

|  |  |  |
| --- | --- | --- |
| **Parameters** | **Values** | **Companies’ view** |
| **UE Speed** | * For spatial domain beam prediction, 3km/h * For time domain beam prediction: 30km/h (baseline), 60km/h (optional) 90km/h (optional), 120km/h (optional) * Other values are not precluded | Supporting companies:   * OPPO, MediaTek, CAICT, Xiaomi, vivo, Futurewei, CATT, CMCC, Fujitsu, ZTE,DCM, Ericsson, Samsung, Lenovo, Qualcomm   Objecting companies: |
| **UE distribution** | * ~~FFS~~ 10 UEs per sector/cell for system performance related KPI evaluation (model inference). * Other values are not precluded * Number of UEs per/sector per cell during data collection (training/testing) is reported by companies if relevant * ~~More UEs per sector/cell for data generation is not precluded.~~ | Supporting companies:   * OPPO, MediaTek, CAICT, vivo, Futurewei, CATT, CMCC, ZTE,DCM, Ericsson, Samsung, Lenovo, Qualcomm, HW/HiSi   Objecting companies: |
| * For spatial domain beam prediction, FFS:   + Option 1: 80% indoor, 20% outdoor as in TR 38.901   + Option 2: 100% outdoor * For time domain prediction: 100% outdoor | Some further discussion is needed |
| **BS Antenna Configuration** | * ~~[~~One panel: (M, N, P, Mg, Ng) = (4, 8, 2, 1, 1), (dV, dH) = (0.5, 0.5) λ as baseline~~]~~ * ~~[Four panels: (M, N, P, Mg, Ng) = (4, 8, 2, 2, 2), (dV, dH) = (0.5, 0.5) λ. (dg,V, dg,H) = (2.0, 4.0) λ as optional]~~ * Other assumptions are not precluded.     Companies to explain TXRU weights mapping.  Companies to explain beam selection.  Companies to explain number of BS beams | Supporting companies:   * OPPO, MediaTek, CAICT, Xiaomi, vivo, Futurewei, CATT, CMCC, Fujitsu, ZTE,DCM,Ericsson, Samsung, Lenovo, HW/HiSi   Objecting companies:   * Qualcomm |
| **UE Antenna Configuration** | * ~~[~~Panel structure: (M,N,P) = (1,4,2)~~]~~   + panels (left, right) with (Mg, Ng) = (1, 2) as baseline * Other assumptions are not precluded     Companies to explain TXRU weights mapping.  Companies to explain beam and panel selection.  Companies to explain number of UE beams | Supporting companies:   * OPPO, MediaTek, CAICT, Xiaomi, vivo, CATT, CMCC, Fujitsu, ZTE,DCM, Ericsson, Samsung, Lenovo, Qualcomm, HW/HiSi   Objecting companies: |
| **Traffic Model** | ~~FFS:~~  For system performance related KPIs (if applicable)   * Option 1: Full buffer (baseline) * Option 2: FTP model (optional)   Other options are not precluded | Supporting companies:   * OPPO, MediaTek, CAICT, Xiaomi, vivo, Futurewei, CATT, CMCC, Fujitsu, ZTE,DCM,Ericsson, Samsung, Lenovo, Qualcomm, HW/HiSi   Objecting companies: |
| **BS Tx Power** | ~~[~~40 dBm~~]~~ | Supporting companies:   * OPPO, MediaTek, CAICT, vivo, Futurewei, CATT, CMCC, Fujitsu, ZTE,DCM,Ericsson, Samsung, Lenovo, HW/HiSi   Objecting companies:   * Qualcomm |

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

|  |  |
| --- | --- |
| Company | Comments |
| FL1 | For UE distribution for spatial domain beam prediction, I would like to collect more views on the two options before pushing to any direction.  For number of UEs and traffic, in FL’s view, it only required for throughput. For L1-RSRP related KPIs, companies can report the number for UEs for data collections and report it if relevant. |
| OPPO | Thanks for the comprehensive summary. We added our view/preference for Proposal 1-1-1a.  For UE distribution of BM-Case1, we believe that both Option 1 (20% outdoor) and Option 2 (100% outdoor) are representative UE distribution in reality. But if we have to choose one option for performance calibration, then we slightly prefer to use Option 2 which is more aligned with BM-Case2 and applicable for the combined spatial + temporal beam prediction. |
| vivo | For UE indoor/outdoor distribution, we support Opt 2 (100% outdoor).    We did some evaluation on the RSRP distribution of Opt 1 as in the figure above. It can be observed that 50% L1-RSRPs in all data are less than -140 dB, while 50% maximum L1-RSRPs per sample are no large than -110dBm. Both of them seem unreasonable due to high penetration loss and more N-Los paths for indoor UEs in FR2. Although such RSRP distribution may not impact on AI model performance in early study if we ignore RSRP quantization, it may lead to the situation that most of the measured or predicted RSRPs are below noise level, and thus difficult to be quantized and reported. Hence, we prefer to adopt the more reasonable UE distribution that is 100% outdoor in spatial domain beam prediction to facilitate performance evaluation and cross-check in RAN1. |
| CATT | For UE distribution, we are open for Option 1 and Option 2. Maybe we can keep both options and encourage companies to report. |
| Fujitsu | To evaluate the system level performance (e.g. throughput), it needs to define the number of TXRU for hybrid beamforming. To simplify the simulation, it can be assumed 1 TXRU per panel and per polarization. In this case the antenna configurations for BS and UE are:  Antenna setup and port layouts at gNB: [4, 8, 2, 1, 1,1,1], (dV, dH) = (0.5, 0.5) λ  Antenna setup and port layouts at gNB: [1,2,1,4,2,1,1], 2 panels (left, right)  Other configuration for the number of TXRU is not excluded. |
| ZTE | For UE distribution, we prefer Option 1 that 80% of UEs are located in indoor and 20% of UEs are located in outdoor as in TR 38.901. Compared with Option 2, the samples collected with Option 1 are more diverse and thus more suitable for performance validation of the spatial domain beam prediction. Besides, even for L1-RSRP related KPIs, more than one UE (10 UEs etc.) per sector per drop can be generated to ease the data collecitons. |
| Ericsson | We support both option 1 and 2 for UE distribution |
| Qualcomm | For BS antenna configuration, for the mentioned BS transmit power (40dBm), the Pout per element should be 22 dBm which is very high, and not realistic. We see that other assumptions are not precluded for BS antenna configuration, but such a high per-element power for baseline is concerning. For the agreed UMa scenario, 64 and 256 dual-pol antenna elements (e.g., for SSB and CSI-RS generation) would be more realistic, as brought up previously. Suggested BS Tx Power is 28 dBm. |
| HW/HiSi | **UE speed:** We would prefer to make one higher speed, 90 km/h, mandatory. The reasons are that this speed is a typical speed in many realistic scenarios and performance curves of the AI approaches are different for different speeds. To obtain more comparable results across companies, we think at least one higher speed, e.g. 90 km/h should also be mandatory to be evaluated.  **UE distribution:** Can accept the updated proposal  **BS antenna configuration:** Can accept the updated proposal  **UE antenna configuration:** Can accept the updated proposal  **Traffic model:** Can accept the updated proposal  **BS TX power:** Ok |

### 1.1.2 Trajectory model for UE mobility

Companies provide views on the three options for UE trajectories:

* Huawei [2]: Proposal 7: For the evaluation of temporal domain beam prediction, Option 4, i.e., random direction straight-line trajectories for randomly dropped UEs, should be considered as the starting point.
* Vivo [3]: Support option #4, i.e. random direction straight-line trajectories, in UE trajectory modelling.
* ZTE [4]: The random direction straight-line trajectories in Option #4 can be adopted for modelling UE mobility, which is simpler than other UE trajectory options and beneficial for model generalization
* Lenovo [11]: Proposal 6: Adopt “Linear trajectory model with random direction change” as the UE trajectory model for temporal beam prediction. (Option 2?)
* Intel [13]:
  + UE trajectories with straight line movement without sharp turns should be considered as a first step for evaluation.
  + The UE trajectory should be sampled at least at the minimum decorrelation distance of the large-scale parameters corresponding to the scenario of evaluation.
* Xiaomi [12]: Option 2 is used in the simulation
* Samsung [17]: Option 2 is used in the simulation
* Nokia [21]:
  + RAN1 further investigates the trajectory model for BM-Case#2, adopting Option #4 as a starting point for further studies.
  + Support RAN1 to further study scenarios/ trajectory model for the BM-Case2.
* MediaTek [22]: we adopt the option-2 [4] user trajectory.
* Qualcomm [24]: The following random trajectory generation process is based on Option 3 in the agreed methods for trajectory generation in RAN1 109e, with some modifications based on Option 4.

Based on the above summary, the following proposals and questions are provided.

#### FL1 (Medium) Question 1-1-2a

**Please share your position on the options:**

|  |  |  |
| --- | --- | --- |
| **Options** | Supporting companies | Objecting companies |
| **Option 2 agreed in RAN 1 #109-e** | Samsung, Lenovo |  |
| **Option 3 agreed in RAN 1 #109-e** |  |  |
| **Option 4 agreed in RAN 1 #109-e** | vivo, ZTE,DCM, HW/HiSi |  |
| **Modified option 3 as in [24]** | Qualcomm |  |

**Please provide your comments, if any.**

|  |  |
| --- | --- |
| Company | Comments |
| FL1 | 5 companies support/use Option 4, and 3(or 4) companies support/use option 2. One company reported modified assumption in the contributions.  In FL’s view, there is no need to down select to one option. However, if some options are not used by any company, it might be better to be removed or further modified. |
| OPPO | In our reading, Option 2 and Option 4 on UE trajectory are not mutually exclusive, i.e. one UE route may satisfy the description in both Option 2 and Option 4 at the same time. Hence, we agree with the assessment of FL that if companies report what UE trajectory model has been used or slightly modified, we don’t have to make a down selection. |
| MediaTek | We think there is no need to downselect one. Companies can report what they used for evaluation. |
| CAICT | We think the three options work and have no strong preference among the 3 options. |
| Xiaomi | Companies can report which option is used for evaluation. |
| vivo | Only option #4 explicitly mentions that the initial UE location as well as the initial moving direction are randomly selected, and there is a clear clarification in the configuration of inter-cell handover/switching and the action after UE trajectory hitting cell boundary. However, option #2 or #3 lacks such definitions. Therefore, we support option #4 considering the completeness of UE trajectory modelling. |
| CATT | We also agree FL’s view, there is no need to down select to one option at least in this meeting. Companies are encouraged to give more simulation results in the future meetings. Then finally, we can remove or modify some options which are not used at all. |
| ZTE | We agree with FL that there is no need to down select to one option. Besides, we prefer Option #4 for modelling UE mobility, which is simpler and beneficial for model generalization. |
| Ericsson | We don’t think there is a need to down select. |
| Samsung | We think no need for downselection for now. |
| Lenovo | Option 2 is simple and is a decent approximation to the real-world UE trajectories. |
| Qualcomm | The modified option 3 that we have proposed tries to address some ambiguities that exist within each of the Options 2,3, and 4, with a concrete model. Options 2,3, and 4 provide the steppingstones to model trajectory, but leave the door open for different interpretations across companies. As an example, Options 2 and 3 do not specify the assumptions regarding initial UE position and direction, and this is only mentioned in Option 4. |
| HW/HiSi | Option 4.  It is a good and simple starting point for the evaluation. It is sufficient to evaluate the aspects of temporal beam prediction and also we think it gives a better foundation to achieve comparable results across companies |

Moreover, there are some discussions on UE orientation together with UE trajectory models:

* Vivo [4]: Slightly prefer option 1b, i.e. randomly per-UE chosen for UE orientation initially but with fixed orientation during SLS, for UE orientation modelling.
  + **Option 1:** UE orientation independent from UE moving trajectory model:
    - **Option 1a:** singular fixed UE orientation for all UE
    - **Option 1b:** randomly per-UE chosen for UE orientation initially, and UE orientation is fixed during SLS.
    - **Option 1c:** randomly per-UE chosen for UE orientation initially, and UE rotation model is followed during SLS.
  + **Option 2:** UE orientation coupled with UE moving trajectory model:
    - **Option 2a:** randomly per-UE chosen for UE orientation initially, and UE's relative orientation with UE's moving direction is fixed during SLS.
    - **Option 2b:** randomly per-UE chosen for UE orientation initially, and UE rotation model is followed for UE's relative orientation with UE's moving direction during SLS.
* Samsung [17]: UE orientation assumption can be reported independently to UE trajectory model by company.
* Qualcomm [24]
  + Proposal 3: Consider the scenario in which the UE orientation changes as a function of UE trajectory.
    - FFS: details of this function

Based on the above summary, the following proposals and questions are provided.

#### FL1 (Low) Question 1-1-3a

**Proposal 1-1-3a**

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

|  |  |
| --- | --- |
| Supporting companies | MediaTek, CAICT, vivo,DCM, Ericsson, Samsung, Lenovo, Qualcomm, HW/HiSi |
| Objecting companies |  |

**Please provide your view Proposal 1-1-3a, if any.**

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

### 1.1.3 Evaluation assumptions with LLS

One company proposed to use LLS for the evaluation of the beam prediction in time domain or for spatial domain beam prediction.

* Samsung [17]:
  + **Data collection:**
    - 8 RBs as baseline, companies can report larger number of RBs
    - First 2 OFDM symbols for PDCCH, and following 12 OFDM symbols for data channel
  + **Channel model:**
    - LOS channel: CDL-D/E extension,
    - NLOS channel: CDL-A/B/C extension,
    - CDL-D extension, DS = 100ns as baseline.
    - Companies explains details of extension methodology considering spatial consistency.
    - Other channel models and DSs are not precluded.

Based on the above proposal, the following proposals and questions are provided.

#### FL1 (Low) Question 1-1-4a

**Proposal 1-1-4a** (updated based on the agreements in RAN 1 #109-e)

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

|  |  |
| --- | --- |
| **Parameter** | **Value** |
| Data allocation | ~~[~~8 RBs~~]~~ as baseline, companies can report larger number of RBs  First 2 OFDM symbols for PDCCH, and following 12 OFDM symbols for data channel |
| Channel model | ~~FFS:~~  LOS channel: CDL-D/E extension, DS = 100ns (CDL-D extension as baseline)  NLOS channel: CDL-A/B/C extension, DS = 100ns  Companies explains details of extension methodology considering spatial consistency  Other channel models are not precluded. |

|  |  |
| --- | --- |
| Supporting companies | DCM,Samsung |
| Objecting companies |  |

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

|  |  |
| --- | --- |
| Company | Comments |
| MediaTek | LLS evaluation can be de-prioritized |
| CATT | LLS evaluation is de-prioritized. |
| NTT DOCOMO | Even though SLS is preferred, we are fine with the assumption for LLS evaluation. |
|  |  |

### 1.1.4 Others

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

* Futurewei [1]: Proposal 1: For AI/ML based spatial beam prediction, to help performance evaluation discussion, companies are encouraged to share simulation details for the dataset generation and provide the average L1-RSRP difference between the ideal L1-RSRP of the Top-1 genie-aided beam and the ideal L1-RSRP of the Top-K genie-aided beams in the training/testing dataset.
* Vivo [3]: It is encouraged for companies to provide publicly accessible dataset and disclose the details for the dataset generation as much as possible for training and validation for cross-check purposes.
* Intel [13]: Spatially consistent large-scale parameter generation should be used for mobility evaluations. Additionally, only spatial consistency model B in [4] can be used for mobility evaluation.
* LGE [18]: As a starting point of UE orientation modelling, it is assumed that UE bearing angle is uniformly distributed on [0, 360] degree, UE downtilt angle is 90 degrees, and UE slant angle is zero degree.
* Ericsson [20]: Observation 1: The agreed simulation scenarios might have heavily skewed beam statistics. AI/ML models can be trained to work well for common beams and ignore uncommon beams. The poor performance of AI/ML models on uncommon beams might not be reflected in average beam prediction statistics. Visualizing the edge percentiles of the L1-RSRP CDF could be one method to illustrate the ability to predict uncommon beams
* NVIDIA [12]: Companies are encouraged to contribute real data to develop and evaluate AI/ML based algorithms for beam management.
* MediaTek [22]: Proposal 3: Study and evaluate the performance of AI/ML beam prediction using the dataset generated by the ray-tracing simulations.
* Ericsson [20]: Observation 9 UE measurement errors may significantly impact ML beam prediction performance and should be considered in realistic evaluations.
* Qualcomm [24]:
  + Proposal 1: Evaluate and identify performance benefits related to beam blockage/failure prediction
  + Proposal 4: For temporal beam prediction, study the impact of incorporating beam prediction quality information (e.g., a measure for prediction confidence such as std of predicted RSRPs) on evaluating the performance of AI/ML model, using the agreed KPIs
  + Proposal 8: For spatial domain beam prediction, study the impact of incorporating beam prediction quality information (e.g., a measure for prediction confidence such as std of predicted RSRPs) on evaluating the performance of AI/ML model, using the agreed KPIs

#### FL1 (Low) Question 1-5-1a

**Please provide your view on the following question:**

1. Whether UE measurement errors shall be considered in the study? If the answer is yes, how to model it?

2. Any other model for dataset generation or other dataset needs to be discussed and agreed on?

|  |  |
| --- | --- |
| Company | Comments |
| FL1 | There is no need to discuss and agree on additional dataset generation for training/testing. If other channel models are used, it can be reported by companies. |
| OPPO | On Question 1: At early stage of AI/ML SI, it seems premature to model the L1-RSRP measurement error for dataset generation for training/testing and for input of inference. Currently, the main goal is to see what performance floor AI/ML model may reach for spatial/temporal beam prediction.  On Question 2: by far it seems the dataset generated by SLS works well and we see no strong need to model other dataset. |
| MediaTek | Agree on FL’s comment. |
| vivo | We are positive to model UE measurement error. Issues like blockage need to be considered. |
| FUTUREWEI | Regarding UE measurement errors, we encourage companies to report how they model it.  As we discussed in our contribution, many details in the simulation will impact the resulting dataset even if companies adopt the agreed-upon parameters, thus, we encourage companies to include the data distribution/analytics of their dataset together with their results, e.g., Top-1 beam distribution and L1-RSRP difference between the ideal L1-RSRP of the Top-1 genie-aided beam and the ideal L1-RSRP of the Top-K genie-aided beams. |
| CATT | Agree with FL. There is no need to discuss additional dataset generation for training/testing. Moreover, we also support companies to share accessible dataset and disclose the details for the dataset generation as much as possible for training and validation for cross-check purposes.  For Question 1, we don’t support to model UE measurement errors in this early stage. |
| Fujitsu | On question 1: it’s too early to model the non-ideal condition (e.g measurement errors) in current stage. |
| ZTE | 1. The UE measurement error can be considered at a later stage if it causes a substantial performance degradation for model inference. Similarly, the quantization error caused by the differential reporting method may also need to be considered. Specifically, if the model inference is performed at the gNB side, a large amount of measured RSRPs need to be quantified and reported from the UE side, which are used as model input. In this case, the associated quantization error may degrade the inference performance and also needs to be considered for evaluation.   2) For cross-checking purposes, publicly accessible dataset can be reported by companies. |
| Ericsson | We think modelling UE measurement errors are needed for the completeness of use case study. It would indicate the prediction performance depends on the UE measurement accuracy. |
| Samsung | 1 Yes. But in later phase. |
| Lenovo | 1.Yes. It’s good to consider errors in UE measurements.  In general, errors in the data set test the “robustness” of the AI/ML model. The model may work very well with clean and accurate data but it’s performance may degrade drastically with noisy data or errors in the data. |
| qualcomm | For AI/ML model performance monitoring, it is important to model UE error and based on the performance try to decide if the model is performing well or not, using a criterion. If the UE is performing beam ID prediction, given the agreed beam accuracy KPIs, the question that needs to be addressed is when we should decide if the UE-side AI/ML model is not performing well. If the UE is performing RSRP prediction, the question is how much the difference between predicted RSRPs is versus actual measured RSRPs for how many prediction instances. These would help in deciding whether to activate/deactivate or switch AI/ML model. |
| HW/HiSI | For Q1: No  For Q2: Spatial consistency procedure should be decided |

## 1.2 Assumptions for inputs/outputs for AI/ML in beam management

### 1.2.1 Clarification on inputs/outputs

Companies have different assumption on the inputs/outputs for AI/ML in beam management. For example:

* Intel [13]
  + BM-Case-1a: Narrow beam measurement-based BS narrow beam prediction with best UE panel and beam selection
  + BM-Case-1b: Wide beam measurement-based BS narrow beam prediction with best UE panel and beam selection
  + BM-Case-9: Joint UE-BS beam-pair-link prediction with best UE panel selection.
* Samsung [17]:
  + Option 1: prediction for Tx beams
  + Option 2: prediction for Rx beams
  + Option 3: prediction for beam pairs
* Nokia [21]:
  + Proposal 8: RAN1 further investigates the comparison between independent Tx beam, Rx beam prediction, and joint Tx-Rx beam pair prediction.
* Qualcomm [24]
  + UE Rx beam prediction: given the best RSRP values for each UE Rx beam at each measured beam management cycle, predict the best UE Rx beams at each predicted beam management cycle
  + gNB Tx beam prediction: given the best RSRP values for gNB Tx beam at each measured beam management cycle, predict the best gNB Tx beams at each predicted beam management cycle

Since the results of different assumptions may lead to very different results that cannot be comparable or may lead to different observations. It may be better to clarify the assumption on the inputs/outputs. The following proposals can be discussed:

#### FL1 (High) Question 1-2-1a

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

* **For AI/ML inputs for both spatial and temporal prediction evaluation, at least the following cases can be further studied:** 
  + **Case A: L1-RSRP of Tx-Rx beam pairs in Set B**
  + **Case B: L1-RSRP of Tx beams in Set B, measured by a “best” Rx beam**
    - **FFS on how to obtain the “best” Rx beam**
  + **Other cases are not precluded**
  + **Note: Other assistance information is not precluded**

|  |  |
| --- | --- |
| Supporting companies | MediaTek, CAICT, Futurewei, DCM, Lenovo, Qualcomm |
| Objecting companies | HW/HiSi |

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

|  |  |
| --- | --- |
| Company | Comments |
| FL1 | Using Tx beam or Tx-Rx beam pair (e.g., Tx beam index or Tx + Rx beam index) as AI inputs needs to be clarified. This might be related to the AI outputs, e.g., Top-K Tx beam(s) or Top-K Tx-Rx beam pair(s).  In FL’s view, this needs further clarification as evaluation assumption. Companies are encouraged to provide other options or modified existing options to make a full list to facilitate the discussion on EVM. |
| OPPO | We are fine to further study both Case A and Case B. However, intuitively the more information input to model (i.e. Case A on Tx-Rx beam pairs), the more prediction can be made (predicted Rx beam when compared with Case B Tx beam only). Additionally, how to determine the best Rx beam may rely on legacy NR beam management mechanism.  One clarification question on bot Case A and Case B is that during inference phase, whether the L1-RSRPs are input to AI/ML model with the same order as that of training phase? If yes, one may consider the Tx and/or Rx beam IDs are implicitly input to model; Otherwise, the Tx and/or Rx beam IDs should be explicitly input to the model in our understanding. Hope this part could be more or less clarified. |
| MediaTek | Prefer to study Case A. For FFS of Case B, can we just do Rx beam sweeping to find the best Rx beam? |
| CAICT | We prefer Case A is used as baseline. |
| Xiaomi | We suggest to add case C   * + **Case C: L1-RSRP of Tx beams in Set B, measured by the same Rx beam**     - **FFS on how to select/configure the same Rx beam** |
| vivo | We agree that Case A and Case B can be further studied and evaluated. But what is the reason to preclude UE Rx beam prediction in this study/evaluation? We suggest to add it like following   * + **Case C: L1-RSRP of Rx beams in Set B, measured by a “best” Tx beam**     - **FFS on how to obtain the “best” Tx beam** |
| CATT | We support Case A. We use L1-RSRP of Tx-Rx beam pairs as inputs in our simulation, i.e., UE use each selected Rx Beam to measure each selected Tx beam and get the L1-RSRP as inputs.  For Case B, how to obtain the “best” Rx beam is needed to be clarified first, before agree on Case B. Similar issue is also needed for Case C proposed by above companies. |
| CMCC | We prefer to prioritize Case A. |
| Fujitsu | Agree with both cases. Our understanding is RSRP and/or beam ID should be input to AI/ML model even though the beam ID may be implicit. |
| ZTE | Both cases can be further studied. Besides, the measured L1-RSRP would always relate to a Tx-Rx beam pair. We suggest the following wording:   * + **Case A: L1-RSRP of Tx~~-Rx beam pairs~~ beams in Set B, measured by multiple Rx beams**   + **Case B: L1-RSRP of Tx beams in Set B, measured by a ~~“best”~~ fixed Rx beam** |
| NTT DOCOMO | Case A is preferred for calibration purpose. |
| Ericsson | We support both cases- |
| Samsung | At least for gNB side inference, there is no need to know the information of Rx beam, i.e., using Tx-Rx beam pair for prediction. Moreover, at least in our understanding, based on the agreements in AI 9.2.3.2 in RAN1#109e, Case B corresponds to BM-Case1 and BM-Case 2 while Case A corresponds to BM-Case9 in spatial domain prediction and not correspond to any BM-Case in temporal domain prediction. Therefore, we have following modification:  **Proposal 1-2-1a:**   * **For AI/ML inputs for both spatial and temporal prediction evaluation, at least the following cases can be further studied:**    + **BM-Case1/2 ~~B~~: L1-RSRP of Tx beams in Set B in both spatial domain prediction and temporal prediction, measured by a “best” Rx beam**     - **FFS on how to obtain the “best” Rx beam**   + **BM-Case9a ~~A~~: L1-RSRP of Tx-Rx beam pairs in Set B in spatial domain prediction**     - **Note: this sub use-case was not agreed as a representative sub-use case in RAN1#109e**   + **Other cases are not precluded**   + **~~Note: Other assistance information is not precluded~~** |
| Lenovo | We support the proposal. Considering and evaluating both (a) Finding Tx-Rx beams jointly (b) finding Tx beam and Rx beams separately, would be interesting. A company may go for either joint Tx-Rx beam finding or finding the Tx separately and, therefore, may choose to opt for a particular input for its AI/ML model. Thus, both Case A and Case B need to be supported and the companies should clearly state what inputs have been used by their AI/ML models. |
| HW/HiSi | Do not agree.  We think Case A is sufficient. In our understanding Case B is included in Case A, since it is not precluded that also in Case A the best RX beam is used.  The bullet that other cases are not precluded, also means that assistance information is not precluded, therefore the note is not needed.  We propose to update the proposal as follows:  **Proposal 1-2-1a (updated):**   * **For AI/ML inputs for both spatial and temporal prediction evaluation, at least the following case~~s~~ can be further studied:**    + **Case A: L1-RSRP of Tx-Rx beam pairs in Set B**   + **~~Case B: L1-RSRP of Tx beams in Set B, measured by a “best” Rx beam~~**     - **~~FFS on how to obtain the “best” Rx beam~~**   + **Other cases are not precluded**   + **~~Note: Other assistance information is not precluded~~** |

### 1.2.2 Number of beams in Set A

Two companies proposed to align the number of beams for Set A and Set B:

* Huawei [2]:
  + Proposal 11: For the evaluation of beam prediction, RAN1 should study multiple sizes of Set A to improve beam management related system performance and overhead KPIs, e.g. to improve the achievable coverage over the legacy baseline.
* Samsung [17]: Align on the number of beams in Set A of beams and Set B of beams for two sub-use AI/ML for BM.
* Nokia [21]:
  + Proposal 1: For BM-Case1, given the current agreed NW antenna configuration, the number of DL Tx beams in Set A should be 32 or 64.
  + Proposal 2: For BM-Case1, RAN1 further study the case of Set A/B are DL Tx and Set B is a subset of Set A.
    - When Set B is a subset of Set A, RAN1 should consider a Set B with a maximum number of DL Tx beams that is ¼ of Set A beams.
* MediaTek [22]:
  + Proposal 6: Study the tradeoff between the beam measurement overhead and prediction accuracy for different number of beams in Set B.

For Set A, the following assumptions are used by companies for Tx beam from gNB:

* 32 Tx beam
  + Futurewei [1]: Horizontal angle = [-60, -42.86, -25.71, -8.57, 8.57, 25.71, 42.86, 60] and Vertical angle = [-45, -20, 5, 30]
  + Vivo [3]: Horizontal angle = [-78.75 -56.25 -33.75 -11.25 11.25 33.75 56.25 78.75], Vertical angle = [22.5 67.5 112.5 157.5]
  + Fujitsu [5]: support 32 transmitting beams (4 beams in vertical and 8 beams in horizonal).
  + OPPO [9]: 32 Tx beams and 4 Rx beams, Zenith angle=[2pi/8, 3pi/8, 4pi/8, 5pi/8]
  + Xiaomi [14]: total beam pairs with 32 TRP Tx beam:
    - 
  + China Telecom [16]: 32 Tx beams: Horizontal angle = [-78.75° -56.25° -33.75° -11.25° 11.25° 33.75° 56.25° 78.75°], Vertical angle = [22.5° 67.5° 112.5° 157.5°]
  + Samsung [17]: There are 4 beams in the vertical direction with 6-degree step, and 8 beams in the horizontal direction within [-60°, +60°] range.
* 64 Tx beam
  + Huawei/HiSi [2]: 64 Tx beams
  + ZTE [4]:64 Tx beams
  + CATT [10]:64 gNB DL Tx beams and 4 UE DL Rx beams
  + NVIDA [12]: Set A consists of 64 beams,
  + Intel [13]: 8x8=64 cross-polarized antenna elements with DFT beams. The 64 gNB antenna elements form 64 narrow beams and 16 virtualized wide beams at gNB.
  + CMCC [19]:
    - The gNB has 64 transmit beams, including 8 horizontal beams and 8 vertical beams, where azimuth angle φi = [-7\*pi/16, -5\*pi/16, -3\*pi/16, -pi/16, pi/16, 3\*pi/16, 5\*pi/16, 7\*pi/16], zenith angle θj = [8\*pi/16, 9\*pi/16, 10\*pi/16, 11\*pi/16, 12\*pi/16, 13\*pi/16, 14\*pi/16, 15\*pi/16].
  + Nokia [21]: Number of BS beams​: 64
  + DoCoMo [26]: TxRU=2, 64 beams are assumed(H(16)\*V(4))
* Others values for Tx beam:
  + Huawei [2]: The AI/ML model infers the Top-K subset from the 256 dense beams to perform beam sweeping at P-2/3. Due to the more precisely selectable beam direction, this achieves better coverage than the legacy exhaustive 64 Tx beam sweeping as long as the AI/ML inferred Top-K beams are accurate.
  + Ericsson [20]: 4x8 Array size: 8 SSB with 30 CSI-RS; 8x16 Array size: 21 SSB with 168 CSI-RS
  + MediaTek [22]: these Set A of beams consist of 24 beams with 2 beam angles along the elevation dimension and 12 beam angles along the horizontal dimension.
  + CEWiT [23]:
    - Case 1: 8 transmitting beams and 8 receiving beams
    - Case2: 8 transmitting beams and 4 receiving beams
    - Case 4: 4 transmit and 4 receive beams.
  + Qualcomm [24]: Temporal: 12 beams (DFT codebook), 6 azimuth beams (22.5 degree separation) and 2 elevation beams (45 degree separation)

The following assumptions are used by companies for Rx beam at UE side:

* 8 Rx beam
  + Futurewei [1]: Horizontal angle = [-60, -20, 20, 60] Vertical angle = [-20, 20]
  + Vivo [3]: 4 Rx beams per panel, Horizontal angle = [-67.5 -22.5 22.5 67.5], Vertical angle = [/]
  + ZTE [4]: 4 Rx beams per panel
  + Fujitsu [5]: configured with 2 panels and total 16 antenna elements which support 8 receiving beams (1 beam in vertical and 4 beams in horizonal for each panel).
  + OPPO [9]: 32 Tx beams and 4 Rx beams per pannel
  + CATT [10]:64 gNB DL Tx beams and 4 UE DL Rx beams
  + Intel [12]: 2 panels and each panel has 1x4 cross-polarized antenna elements
  + Xiaomi [14]: 8 UE Rx beam with 2 panels.
  + China Telecom [16]: 8 Rx beams: Horizontal angle = [-67.5° -22.5° 22.5° 67.5°], Vertical angle = [45° 135°]
  + CMCC [19]: UE has 4 transmit beams, including 1 horizontal beam and 4 vertical beams, where azimuth angle φi = [-3\*pi/8, -pi/8, pi/8, 3\*pi/8], zenith angle θj = pi/2.
  + Nokia [21]: number of UE beams: 8

Based on the above summary, the following proposals can be discussed:

#### FL1 (High) Question 1-2-2a

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

* **For the evaluation of both temporal and spatial domain prediction, Set A consists of 32 or 64 Tx beams. Other values are not precluded and can be reported by companies.**

|  |  |
| --- | --- |
| Supporting companies | MediaTek, CAICT, Futurewei, fujitsu, ZTE, DCM, Samsung, Lenovo, Qualcomm |
| Objecting companies | HW/HiSi |

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

* **For the evaluation of both temporal and spatial domain prediction, 8 Rx beams are used at UE side. Other values are not precluded and can be reported by companies.**

|  |  |
| --- | --- |
| Supporting companies | MediaTek, CAICT, FUJITSU, DCM, Lenovo, Qualcomm |
| Objecting companies |  |

**Please provide your view Proposal 1-2-2a and Proposal 1-2-3a, if any.**

|  |  |
| --- | --- |
| Company | Comments |
| FL1 | In order to draw observations, it is better to align the number of beams. |
| OPPO | We are okay with 32 or 64 Tx beams and 8 Rx beams.  But the formulation on Proposal 1-2-2a (Set A including Tx beams only) and Proposal 1-2-3a (Set A including no Rx beams) seems asymmetric. Given the possibility that Tx-Rx beam pairs as input, we suggest to slightly rewording Proposal 1-2-2a as below.  **Proposal 1-2-2a:**   * **For the evaluation of both temporal and spatial domain prediction, ~~Set A consists of~~ 32 or 64 Tx beams are used at NW side. Other values are not precluded and can be reported by companies.** |
| Xiaomi | We share same confusion as OPPO. From the proposal 1-2-2a and 1-2-3a, it seems set A only consists of Tx beam, and not consist of Rx beam.  We are ok with the version updated by OPPO. |
| vivo | We are okay in general, but it needs to be noted that other values should also be considered when discussing issues like generalization, e.g.,  **Note: other values shall be considered for issues like generalization** |
| CATT | We share the same view as OPPO. We are fine with 32 or 64 Tx beams and 8 Rx beams. Proposal 1-2-2a needs to be updated as OPPO proposed. |
| CMCC | OPPO’s update is fine to us. |
| Fujitsu | Agree with OPPO modification on proposal 1-2-2a. |
| ZTE | We support FL’s proposal. Note that only beams in set B need to be measured and the beams in set A may not be transmitted by gNB. |
| NTT DOCOMO | We support the proposal.  Please remove the ‘Set A consists of’ in Proposal 1-2-2a as OPPO indicated. |
| Ericsson | We support to agree on a common value for number of assumed beams. We are fine with 32 or 64 TX beams, 8 RX beams. |
| Samsung | Regarding RX beams, it’s up to the number of RX panel. Therefore, we suggest following revision:  **Proposal 1-2-3a:**   * **For the evaluation of both temporal and spatial domain prediction, ~~8~~4 Rx beams per UE panel are used at UE side. Other values are not precluded and can be reported by companies.** |
| Qualcomm | We believe there is no strong justification to limit the codebook size at this point and this should be left open for the companies to explore and analyze the comparative performance of different codebook sizes. However, agree with the above proposals to make progress. OK with OPPO’s update. |
| HW/HiSi | **Proposal 1-2-2a:**  The useful number of beams in set A is different depending on whether the use case is for spatial BM or temporal BM. For spatial domain BM we have shown that AI/ML based approach considered 256 beams Set A gives 1.1 dB gain over exhaustively sweeping 64 beams considered 64 beams Set A but with 1/4 inference overhead. For temporal domain BM, the maximum number of beams in Set A can be smaller.  It should be noted that the size of set A has no impact on inference, just on the training overhead. Everything else, is independent from Set A (same AI model, same input, same overhead, same top-K, etc).  **Therefore, we think that no agreement on the size of set A is needed. But if the group really wants to decide on the size of Set A, we suggest to update the proposal with one of the two following alternatives**  Alt A:   * **For the evaluation of both temporal and spatial domain prediction, Set A consists of ~~32 or~~ 64 Tx beams. Other values are not precluded and can be reported by companies.**   Al B:   * **For the evaluation of ~~both temporal and~~ spatial domain prediction, Set A consists of 32 or 64 or 256 Tx beams.** * **For the evaluation of temporal domain prediction, Set A consists of 32 or 64 or 256 Tx beams.** * **Other values are not precluded and can be reported by companies.**   **Proposal 1-2-3a:**   * **For the evaluation of both temporal and spatial domain prediction, 8 Rx beams are used at UE side. Other values are not precluded and can be reported by companies.** |

### 1.2.3 Selection of Set B of beams

There were some discussions on the selection or design of Set B of beams:

* Futurewei [1]
  + Fixed pattern: (even-space sampling)
  + Random pattern
  + Pre-set pattern:
    - In this option, we pre-defined a set of N (N = 5) different beam patterns, each with M selected beam pairs (M ∈ {4, 8, 12, 16, 20, 24, 28, 32}), then one of them will be randomly chosen as input for each sample.
* Vivo [3]:
  + **Set 1:** Fixed subset with continuous beams
  + **Set 2:** Fixed subset which is randomly selected
  + **Set 3:** Well-designed subset
  + **Set 4:** Best fixed subset
* ZTE [4]:
  + The input of the AI model has two pattern options: fixed beam pattern and random beam pattern
* China Telecom [15]:
  + Using a fixed beam selection pattern for AI/ML model training and testing can be a starting point for spatial beam prediction performance evaluation.
* CMCC [19]
  + Three fixed beam measurement patterns are considered, i.e. 4 transmit beams ×4 reception beams, 8 transmit beams ×4 reception beams, 12 transmit beams ×4 reception beams.
* Nokia [21]
  + Proposal 4: For BM-Case1, RAN1 further study Set B to be a fixed pattern.
  + Proposal 5: For BM-Case1 model inference applies at the NW side, with DL Tx beams considered for Set A and Set B, the training a model with random Set B is not needed.
  + Proposal 7: For BM-Case1, RAN1 further study the case of Set A/B are DL Tx and Set B/Set A are different.
    - Set B is a wide beam codebook and Set A is a refined beam codebook
    - Advance Set B designs are needed to provide sufficient refined beam prediction performance.
  + Proposal 11: For BM-Case2, support RAN1 to further study the following methods:
    - Method 1: Set B is a fixed subset of Set A
    - Method 2: Set B is a variable subset of Set A
    - Method 3: Set B is the same as Set A
    - Methods 1 or 3 + Assistance Info: ML model input consists of L1-RSRP measurement based on Set B and assistance information
* MediaTek [22]:
  + Proposal 9: For AI/ML-based spatial domain beam prediction evaluation, study the subset selection if Set B is a subset of Set A.
* Qualcomm [24]:
  + Method 1A: pick best beam from Set ( in Figure 11) and Set ( in Figure 11) at UE & gNB, respectively, using AI/ML model. As AI/ML inference is being done at the UE side, UE needs to feedback best beam index from beam set to gNB.
  + Method 1B: gNB uses best beam from its codebook (Set ) and UE uses best beam from Set ( in Figure 12), using AI/ML model
* MediaTek [22]:
  + Observation 5: The spatial beam prediction by using multi-arm beam design in Set B performs better than using subset beam design in Set B.
  + Proposal 10: Study and evaluate a more comprehensive Set B design, including joint designing the number of beams in Set B and their beam shape for spatial beam prediction.
* DoCoMo [26]: The mapping method of wide/sparse beam as input into AI/ML model should be investigated.

Some observations with fixed beam or random beams:

* Futurewei [1]
  + Observation 1: We observe when using the same amount of training samples and same number of input beam measurements, fixed input beam pattern achieved better performance compared with random input beam patterns and pre-configured input beam patterns.
* Vivo [3]:
  + Observation 2: Fixed beam subset in Set B can have good performance in ideal scenarios but it lacks flexibility. Issues like blockage and inter-cell interference can bring negative impact on the performance of fixed subset.
* ZTE [4]:
  + Observation 4: The spatial domain beam prediction with the fixed beam pattern obtains a better performance than that of the random beam pattern.
* CATT [10]:
  + Observation 1: Beam prediction accuracy with fixed pattern has better performance than random pattern.
  + Observation 9: For BM-Case1\_(DL Tx) model inference in UE side, training model with random Set B may reduce model switching/indication/ transferring overhead for UE. But the benefit of BM-Case1 (DL Tx) model inference on the UE side is not yet clear.
* MediaTek [22]:
  + Observation 4: The selection of beams in Set B will affect the prediction accuracy of the AI/ML-based spatial beam prediction.
* Nokia [21]
  + Observation 2: For BM-Case1, a “sparse” Set B, or a random Set B pattern design, may cause throughput loss, especially for the cell-edge UE.
  + Observation 5: For BM-Case1, the ML model using as input only RSRP measurements has performances that reduce significantly changing the number of RSRP measurements from 8 to 4, i.e. further down sampling Set A, from a ratio of ¼ to a ratio of 1/8.

Moreover, the assumptions on number of beams (pair) in Set B used by each company are summarized:

* Futurewei [1]: 4, 8, 12, 16, 20, 24, 28, 32
* ZTE [4]: 16 fixed/random beams with uniform sampling
* Fujitus [5]: 16 transmitting beams and 4 receiving beams (total 64 beam pairs)
* CATT [10]: The number of beam pairs in Set B is 32, which includes selected 8 gNB DL Tx beams and 4 UE DL Rx beams.
* NVDIA [12]: Set B consists of 16 beams
* Intel [13]: For example, 8, 10, 12, and 16 beams are measured in case-1a and case-1b. 16 and 32 beams are measured in case-9.
* Xiaomi [14]: The number of input beam pairs’ RSRP is 64
* Samsung [17]: 4 or 8
* Nokia [21]: 4, 8, 16, 32

#### FL1 (High) Question 1-2-3a

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

* **Further study the follow options on the selection of Set B of beams (pairs),** 
  + **Option 1: Set B is a fixed subset** 
    - **FFS on the pattern of Set B**
  + **Option 2: Set B is a variable subset (i.e. different beam (pairs) patterns in each report during training and/or inference)** 
    - **FFS on the details**
  + **Other options are not precluded.**
  + **FFS on the number of beam (pairs) in Set B**

|  |  |
| --- | --- |
| Supporting companies | OPPO, CAICT, CATT, CMCC, FUJITSU, ZTE, DCM, Ericsson, Samsung, Lenovo, Qualcomm |
| Objecting companies |  |

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

|  |  |
| --- | --- |
| Company | Comments |
| FL1 | It is quite important to clarify on the selection of Set B. Encourage companies to provide modified wording for the descriptions of each option or add other options. |
| MediaTek | 1. Support to study both Option1 and Option2. For Option1, companies are encouraged to provide the subset pattern of beams used in Set B or how to select Set B from Set A.  2. In this proposal, the discussions are only under the assumption that Set B is a subset of Set A. However, in agenda 9.2.3.2, we are having a discussion to study the case when Set B has a different design of Set A (Proposal 2.2.2-1, Alt.1). For example, in our evaluation result [22], we observed that by using the multi-arm beam design in Set B, the AI/ML model can perform better compared to Set B is a subset of Set A, without any further measurement on narrow beams. Therefore, we want to propose to study and evaluate a more comprehensive Set B design, including joint designing the number of beams in Set B and their beam shape (Proposal 10 [22]). |
| Xiaomi | For option 2, it is better to make it clear that “**Option 2: Set B is a variable subset with fix or variable number of beams (pairs)**”. |
| vivo | We are okay to consider these two but the description of Opt 1 and Opt 2 needs to be accurate.  For Opt 1, we understand this “fixed” means to fix it in both training and inference, hence we suggest the following change.  **Option 1: Set B is a fixed subset across training and inference**  For Opt 2, it’s too early to discuss what to report or how to report. Hence we suggest the following change.  **Option 2: Set B is a variable subset (i.e. different beam (pairs) patterns in each ~~report~~ measurement during training and/or inference)** |
| FUTUREWEI | Option 2 needs to be clarified regarding whether variable subset means each sample has different number of beam pairs in Set B, or the number of beam pairs in Set B is the same across all samples while their patterns can be different. |
| CATT | For Option 2, we prefer the number of beam pairs in Set B is the same across all samples while their patterns can be different. |
| CMCC | Support to study Option1 and Option2. For option 2, when Set B is a variable subset, it can be used for generalization evaluation. |
| Fujitsu | Option 1 can be considered as the baseline. |
| ZTE | Both options can be considered. According to evaluation results of companies, Option 1 has potential to obtain a better beam prediction performance than that of the variable subset in Option 2. Nevertheless, Option 2 with variable subset can also be studied due to its advantages in configuration flexibility and model generalization. The variable subset may includes different numbers of beams with random beam pattern. |
| Ericsson | Option 1 should be mandatory, option 2 could be optional. |
| Samsung | We also observe that fixed set has better performance.  Before agreeing on this, we would like to discuss the motivation to support random/variable set |
| Lenovo | Support Option 2. When we need to determine the best beam from Set A, the AI/ML model should be allowed to choose what beams it needs to measure for each instance of beam management. What should be of concern is the number of beams measured by the AI/ML model, but not what beams are measured. Thus, only the cardinality of Set B is important (and it determines the amount of overhead and latency of the beam management process) but the actual elements of Set B should be left for the AI/ML model to select as per its requirement, in a dynamic manner. |
| Qualcomm | We believe the motivation and evaluation results related to option 2 should be discussed further to justify this option. |
| HW/HiSi | Do not support. We are wondering how much this proposal would bring the group forward. According to our understanding both options had been studied until this meeting.  We observe that to this meeting a clear majority found a fixed pattern achieved better performance, and a fixed pattern also is simpler. Therefore we think the fixed pattern should be prioritized. But of course, other pattern can still be studied.  We propose to update the proposal as follows:  **Proposal 1-2-4a:**   * **Prioritize ~~F~~further study of the following option~~s~~ on the selection of Set B of beams (pairs),**    + **Option 1: Set B is a fixed subset**      - **FFS on the pattern of Set B**   + **~~Option 2: Set B is a variable subset (i.e. different beam (pairs) patterns in each report during training and/or inference)~~**      - **~~FFS on the details~~**   + **Other options, such as Set B is a variable subset, are not precluded.**   + **FFS on the number of beam (pairs) in Set B** |

### 1.2.4 Assumption of time domain information

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

* Huawei [2]
  + Proposal 12: To reduce the overhead of temporal beam prediction, sparse beam sweeping should be used to generate the input to the AI/ML model, where Set B should be a subset of Set A or a different set from Set A but with smaller size.
* vivo [3]
  + Observation 6: For BM Case 1, compared with AI based P1 procedure,
    - P2 procedure assuming the best Rx beam can achieve better performance.
    - P2 procedure assuming the 2nd best Rx beam does not have clear loss in terms of RSRP difference but has some loss on beam accuracy.
    - If the Rx beam assumptions of training and inference are different, performance loss can be observed for P2.
  + Observation 9:
    - For BM-case2, compared with non-AI 2-step scheme, AI based 2-step scheme improves beam prediction accuracy and reduces average L1-RSRP difference significantly.
* Rakuten Mobile Inc. [8]:
  + Proposal 1: Consider a two-step beam management procedure where existing beam management mechanism is used to choose the best beam from a set of beam recommendations from the AI/ML engine.
* Samsung [17]:
  + Align the assumption on SSB/CSI-RS patterns in time domain at least for temporal beam prediction.
  + The input/output for AI/ML needs to be clarified, together with the assumption on beam management procedure and RS time domain pattern for measurement.
* Ericsson [20]: For example, if Set B is SSB beams, we have the following steps:
  + P1. Measure the SSB beams
  + P1\*. Predict a set of *K* CSI-RS beams
  + P2. Measure the *K* CSI-RS beams
* Nokia [21]:
  + Proposal 12: For BM-Case2, with Set B is a subset of Set A, measurement instances K and prediction instances F shall be carefully investigated prior supporting the sub-use case.
  + For each UE, we collected the RSRP values measured at the UE assuming UE applies the optimal Rx beam, and with the gNB sweeping all the Tx beams over successive synchronization signal blocks (SSBs) and repeating the same operation for different Rx beams.

Other than the assumption of management procedure and RS patterns, the time domain assumption may impact on the temporal beam prediction. For example, in TS 38.214, if configured by higher layer, the measurements for computing L1-RSRP reports based on certain resource, as:

*If the higher layer parameter timeRestrictionForChannelMeasurements in CSI-ReportConfig is set to "Configured", the UE shall derive the channel measurements for computing L1-RSRP reported in uplink slot n based on only the most recent, no later than the CSI reference resource, occasion of SS/PBCH or NZP CSI-RS (defined in [4, TS 38.211]) associated with the CSI resource setting.*

Therefore, it is better to align the time domain assumptions at least for temporal beam prediction. Therefore, the following proposals can be discussed:

#### FL1 (Medium) Question 1-2-4a

**Proposal 1-2-5a:**

* **At least for temporal beam prediction, further study on the time domain assumptions, at least including:**
  + **RS pattern for L1-RSRP measurement**
  + **Beam management procedure**

|  |  |
| --- | --- |
| Supporting companies | MediaTek, DCM, Samsung(With proposed updates), Lenovo, Qualcomm, HW/HiSi |
| Objecting companies |  |

**Please provide your view Proposal 1-2-5a, if any.**

|  |  |
| --- | --- |
| Company | Comments |
| MediaTek | 1. Suggest changing the term “RS pattern” to “RS pattern in time domain” or “RS time configuration”.  2. What does “beam management procedure” exactly mean in the second sub-bullet? |
| Xiaomi | It is better to clarify the meaning of RS pattern and the beam management procedure. |
| vivo | We seek for clarification on RS pattern. It should include issues like construction of set B if it is a subset of set A, and time gap between adjacent occasions for measurement? If so, we’d better clarify this, like following   * + **RS pattern for L1-RSRP measurement** * Include issues like construction of set B if it is a subset of set A, and time gap between adjacent occasions for measurement |
| CATT | Need further clarification on RS pattern. In our understanding, RS pattern can re-used the legacy one. For beam management procedure, is that means for beam measurement reporting when AI/ML model deployed in NW side? |
| NTT DOCOMO | We prefer to make “*timeRestrictionForChannelMeasurement*s in *CSI-ReportConfig* is set to "Configured"” as baseline for the simplicity.  If the L1-RSRP for model inputs is calculated based on CSI-RS resources at several time instances (e.g., *timeRestrictionForChannelMeasurements* is not set to “Configured”) in the simulation, how L1-RSRP is calculated should be clarified by each company for calibration purposes. |
| Ericsson | Share the view around unclarities for RS-pattern. Needs to be clarified |
| Samsung | We suggest to change RS pattern to “RS time domain pattern”. We think that, the time domain density will impact on the performance of temporal beam prediction.  Moreover, beam management procedure can be discussed together with Rx beam or Tx-Rx beam pair |

### 1.2.5 Assistance information

Assistance information were discussed and some observations are summarized:

* Vivo [3]:
  + Observation 3: For random subset selection, i.e., training and inference use different beam subsets in Set B, to include Tx/Rx beam ID or angle into the AI model is helpful to reduce performance loss.
  + Observation 4: To restrict the selection of random subset from the best X beam subsets can improve the performance of BM Case 1 prediction. Such semi-random selection with Tx/Rx beam angle information as input barely suffers performance loss compared with the best beam subset.
  + Observation 5: To input the expected Tx and/or Rx beam information in the AI model can enable the utilization of a trained AI model to different numbers of Tx or Rx beams with marginal performance loss.
* ZTE [4]:
  + Observation 5: If the associated beam ID is used as an assistance information for the AI input, the beam prediction accuracy for the random beam pattern can be greatly improved.
* Ericsson [20]:
  + Best beam statistics can be highly biased (e.g., beams pointing near the horizon). Therefore, mean-user KPIs (e.g., average probability of selecting the best beam) can be insensitive to tail events (e.g., UEs in uncommon locations).
  + There is a strong correlation between UE location and best beam direction, and hence UE location can be considered as assistance information.
  + Observation 8 UE location as assistance information can substantially improve prediction performance for outdoor UEs.
* Nokia [21]
  + Observation 6: For BM-Case1, when the ML model use the UE angle as the assistance information, it has a better performance than all the other variants.
  + Observation 7: For BM-Case1, the ML model using as input RSRP measurements and UE Position has performances that outweigh the performance of the ML model using only RSRP.
  + Observation 8: For BM-Case1, using assistance information like Beam Angle and Beam ID related to the measured beams may not significantly improve the performance of the ML model using as input only RSRP with a fixed pattern.
  + Proposal 3: For BM-Case1, RAN1 further study the use of assistance information at the ML model input. The following assistance information can be prioritized:
    - the beam angle and/or the beam boresight direction for the measured DL Tx beams from NW to UE.
    - the UE position information.
    - the UE’s angle relative to a panel array of the gNB
  + Proposal 13: For BM-Case2, RAN1 further verify whether there is any use of using assistance information at the input of the ML model. The following assistance information can be further considered:
    - the UE position information.
* MediaTek [22]:
  + Proposal 7: For AI/ML-based spatial domain beam prediction evaluation, adopt the RSRP of beams in Set B as the AI/ML model inputs. Additional information to the input of AI/ML model is not excluded.
* Qualcomm [24]
  + (Temporal) we assume signalling of assistance information and rely on channel impulse responses (CIRs) of top- beam pairs as inputs to the AI/ML model.
  + (Spatial domain): As mentioned in the beginning of this section, we consider UE-side AI/ML models and assume signaling of assistance information from gNB to UE. The assistance information includes beam boresight directions of beams from Set and , and also location vector of gNB panel antenna elements, from gNB to UE. Please note that this assistance information is used for *both* Method 1A and Method 1B.

#### FL1 (Low) Question 1-2-5a

**Proposal 1-2-6a:**

* **At least for spatial domain prediction, other than beam (pair) ID related information, the following assistant information can be further studied:** 
  + **Tx and/or Rx beam angle**
  + **Tx and/or Rx beam boresight direction**
  + **UE position information**
  + **FFS on other assistant information**

|  |  |
| --- | --- |
| Supporting companies | MediaTek, Futurewei, FUJITSU,DCM, Lenovo, Qualcomm |
| Objecting companies |  |

**Please provide your view Proposal 1-2-6a, if any.**

|  |  |
| --- | --- |
| Company | Comments |
| MediaTek | For each assistance information, companies should consider and explain how to obtain the information at the host where the AI/ML model is inferenced. |
| vivo | We suggest to add “**Expected Tx and/or Rx beam information**” as a bullet. This is used to control what the output beams are. We think it is critical to enable the utilization of a trained AI model to different numbers of Tx or Rx beams with marginal performance loss. |
| CATT | We are fine with this proposal. Moreover, the beam (pair) ID is needed to further clarify, whether the beam (pair) ID is implicit or explicit and how to input. The benefit of beam (pair) ID can be certified as well. |
| Fujitsu | It’s encouraged for companies to show the gain with the assistance information when the evaluation results are summited. |
| NTT DOCOMO | We prefer to add the beam shape as well, such as relative power of each beam at each angle. |
| Ericsson | Supporitve of UE position. We don’t think it is useful to consider angle and direction. It assumes using a DFT-based precoding. This might not be applicable for real deployment scenarios. |
| Samsung | Before agreeing on the list, we prefer to discuss the feasibility to obtain the information. For example, for Tx beam angle, we are not sure gNB can know it. For UE position, it might be hard to verify the accuracy of the information. We suggest to focus on L1-RSRP reports only in this first stage. |
| Lenovo | UE position information could be a very useful assistance information. Open to consider other assistance information as well. When the proposed AI/ML model uses any kind of assistance information, the efforts required (such as the overhead, latency, computational complexity, any additional hardware/sensors required at the UE/gNB etc.) should clearly be stated and accounted for while assessing the performance of the AI/ML model. |
| Qualcomm | What is the difference between first and second bullet? Is the first bullet meant to be ‘beam pointing angle’? In this case it is equivalent to beam boresight direction. |
| HW/HiSI | We would like to have clarified if beam pair ID related information already has been agreed, what is meant specifically with that? |

### 1.2.6 Others

Some other input/output related discussion:

* Huawei [2]
  + Proposal 2: The evaluation for beam prediction should take one-sided AI/ML model into consideration, including the Network-side AI/ML model and UE-side AI/ML model.
  + FL1: Can be discussed in 9.2.3.2
* MediaTek [22]:
  + Proposal 8: Adopt one of the following as the output of AI/ML model: (i) beam index of highest RSRP Set A of beams. (ii) RSRPs of all the Set A of beams.
  + FL1: Can be discussed in 9.2.3.2

## 1.3 Generalization

Generalization is one of the important aspects to verify the performance of AI/ML model. The follow discussions/proposals were found:

* Huawei/HiSi[2]:
  + Proposal 3: To verify the generalization of AI/ML models on AI/ML-based beam management in both spatial and temporal domains, the following cases to construct the training dataset and testing dataset should be considered:
    - Case 1: The AI/ML model is trained based on training dataset from one Scenario#A/Configuration#A, and then the AI/ML model is tested on dataset from the same Scenario#A/Configuration#A
    - Case 2: The AI/ML model is trained based on training dataset from one Scenario#A/Configuration#A, and then the AI/ML model is tested on dataset from a different Scenario#B/Configuration#B
    - Case 3: The AI/ML model is trained based on training dataset constructed by mixing datasets from multiple scenarios including Scenario#A/Configuration#A and Scenario#B/Configuration#B, and then the AI/ML model is tested on dataset from a single Scenario#A/Configuration#A or Scenario#B/Configuration#B from the multiple scenarios
    - Case 4: The AI/ML model is trained based on training dataset from one Scenario#A/Configuration#A, and then the AI/ML model is fine-tuned based on the fine-tuning dataset from a different Scenario#B/Configuration#B. After that, the AI/ML model is tested on dataset subject to the same Scenario#B/Configuration#B as the fine-tuning dataset
  + Proposal 5: For verifying the AI/ML model generalization for spatial domain beam prediction, the scenarios/configurations for performing the inference for the AI/ML model include at least:
    - Various channel types, including UMa, UMi, InH
    - Various indoor/outdoor ratios, including 10:0, 8:2, 5:5, 2:8, 0:10
    - Various numbers of beams in Set A/Set B
  + Proposal 8: For verifying the AI/ML model generalization for temporal domain beam prediction, the scenarios/configurations for performing the inference for the AI/ML model include at least:
    - Various channel types, including UMa, UMi, InH
    - Various indoor/outdoor ratios, including 10:0, 8:2, 5:5, 2:8, 0:10
    - Various UE speeds (e.g., 30km/h, 60km/h, 90km/h)
    - Various types of trajectories
    - Various numbers of beams in Set A/Set B
* Vivo [3]
  + Support to define generalization performance KPI.
  + To study and evaluate generalization, at least the aspects including different scenarios, different UE speeds, different number of Tx beams and Rx beams, and different gNB/UE antenna configurations, should be prioritized.
  + For evaluation of generalization performance, support to evaluate KPIs for a separately generated testing dataset generation method with 1 target parameter difference. Multiple target parameters can also be verified in further study.
* Lenovo [11]
  + Proposal 1: For testing the generalizability of an AI/ML model for beam management, consider relevant network conditions/scenarios/parameter values. The full list of such network conditions/scenarios/parameter values need to be discussed and decided.
  + Proposal 2: Generalizability of a proposed AI/ML model for beam management can be evaluated by computing all the KPIs under each of the different network conditions/scenarios/parameter values.
* CAICT [15]
  + Proposal 1: the definition of generalization of AI model for BM should consider training set and testing/inference set in the same scenario(s) with different configuration.
* Samsung [17]
  + For UE side inference, different scenarios and cell/gNB specific configurations/parameters are considered
  + For gNB side inference, different UE specific configurations/parameters are considered
* CMCC [19]
  + Generalization capability of AI model under different scenarios and beam measurement configurations (at least including different number and combination of measured beam pairs in set B) needs further evaluation.
* Ericsson [20]
  + Further study AI/ML model generalization in beam management, evaluating the inference performance of beam prediction under multiple different scenarios/configurations.
  + It is too early to define the exact scenario for testing model generalizations at this stage of the study item. Companies are encouraged to propose scenario alternatives
  + When reporting results, the proponents should describe the type of generalizability targeted by the experiment (e.g., generalize AI/ML beam prediction model over different sites/cells, carrier frequencies, antenna configurations, reference signal configurations). The proponents should explain how their training, validation, and testing procedure supports the claimed generalizability.
* Nokia [21]
  + Observation 10: For BM-Case1, the Set A/B model generalization issue can be addressed with a training model based on an oversampled Set C that satisfies Set B∈Set A∈Set C for any given Set A/B.
  + Proposal 6: For BM-Case1, RAN1 further study the model generalization issue considering the Set A/B dimensions.
* Qualcomm [24]
  + Proposal 2: Consider the following categorizations for definition of scenarios/configurations for evaluating the generalization capability of AI/ML models for temporal beam prediction:
    - Inter-site (heterogeneous): train AI/ML model on a first set of deployment type(s) and test it on a second (unseen) deployment type.
    - Inter-site (homogeneous): train on a first set of site(s) of a given deployment type and test it on a second (unseen) site of that same deployment type.
    - Intra-site: train AI/ML model for a given site and test it on unseen variations within that same site.
    - Across configurations: train AI/ML model on a first set of configuration(s) and test on a second configuration
  + Proposal 6: Consider the following categorizations for definition of scenarios/configurations for evaluating the generalization capability of AI/ML models for spatial domain beam prediction:
    - Inter-site (heterogeneous): train AI/ML model on a first set of deployment type(s) and test it on a second (unseen) deployment type.
    - Inter-site (homogeneous): train on a first set of site(s) of a given deployment type and test it on a second (unseen) site of that same deployment type.
    - Intra-site: train AI/ML model for a given site and test it on unseen variations within that same site.
    - Across configurations: train AI/ML model on a first set of configuration(s) and test on a second configuration
* Apple [25]
  + Proposal 2: For generalization, discuss aspects related to analog beam design, antenna configurations including M/N, and antenna spacing.

Observations:

* Vivo [3]
  + Observation 7: If the beam shape for training and beam shape for inference are different, performance loss can be observed.
* Ericsson [20]
  + Observation 10 Initial evaluations indicates that a model trained in one cell is found to generalize to another cell with identical antenna configuration
* Nokia [21]
  + Observation 11: For BM Case1, the training model with a fixed Set B pattern will have poor beam prediction performance if the testing Set B does not match with the training Set B.
  + Observation 12: For BM-Case1, training the model with random Set B is possible to provide beam prediction performance close to the optimal case – training and testing on the same Set B.

Based on the above analysis, the following proposals can be discussed:

#### FL1 (High) Question 1-3-1a

**Please provide your views on the scenarios/configurations for evaluating generalization performance:**

|  |  |  |
| --- | --- | --- |
| Company | Scenarios/configurations | Comments |
| FL1 | Companies are encouraged to provide a list of scenarios/configuration | Companies are encouraged to provide values of the proposed of scenarios/configuration, and applicable case (e.g., BM-Case1 or BM-Case2), and training/inference location. |
| MediaTek | 1.Two cells using different gNB Set B, including subset pattern and size of beams.  2.Multiple groups of UEs with different UE Set B, including subset pattern and size of beams.  3. Two cells using same gNB Set B, but channel conditions are different (e.g. different buildings, road conditions)  4.LOS and NLOS scenarios  5.Indoor and Outdoor scenarios | 1. Set B is a subset of Set A. Applicable case: for BM-Case1. Model is inference at the UE side.  2. Set B is a subset of Set A. Applicable case: for BM-Case1. Model is inference at the gNB side  3. Set B can be same or different design of Set A. Applicable case: for both BM-Case1 and BM-Case 2. Model can be inferenced at gNB or UE side.  4. Set B can be a valuable design of Set A. Applicable case: for both BM-Case1 and BM-Case 2. Model can be inferenced at gNB or UE side.  5. Set B can be same or different design of Set A.. Applicable case: for both BM-Case1 and BM-Case 2. Model can be inferenced at gNB or UE side. |
| CAICT | Based on the agreements in last meeting, we think the evaluation of generalization of AI model is based on training set and testing/inference set in Uma with different configuration | For BM-Case1, different UE numbers per drop could be considered for training and inference dataset to verify the generalization of model with the agreed simulation assumptions.  For BM-Case2, different UE speeds setting could be used for generalization verification. |
| vivo | Consider the following different aspects in training and inference   * Different scenarios, * Different UE speeds, * Different number of Tx beams and Rx beams,   Different gNB/UE antenna configurations | * Scenarios like Indoor v.s. Uma/Umi * Different UE speeds can be 30, 60, 90, etc, especially for BM Case 2 * Different number of beams can be   + Different number of predicted Rx beams as output, e.g., Multiple Rx beams v.s. 1 Rx beam   + Different number of predicted Tx beams as output, e.g., 32, 64, 128 Tx beams   Different antenna configurations can be different numbers of antennas which form different beam shapes |
| Futurewei |  | We suggest discussing various model generalization aspects first.  In general, the following aspects can be considered:   * + - 1. Scenario-based deployment: verify model generalization using data unseen during the training phase but is generated from the same distribution (scenario/configuration) as the training dataset.       2. Generalized deployment aspect A: verify model generalization using test data generated from a different distribution (scenario/configuration) from the training dataset and no labelled test data is seen or being used during the training phase.       3. Generalized deployment aspect B: verify model generalization using test data generated from a different distribution (scenario/configuration) from the training dataset while a small portion of the new labelled test data may be seen or used during the training phase.       4. Generalized deployment aspect C: verify model generalization using test data generated from a different distribution (scenario/configuration) while sufficient (or full) labelled test data is seen or being used during the training phase.   For BM-Case 1, the following scenarios/configurations can be considered as starting point:  Scenarios: UMa/UMi/InH  UE distributions: indoor, outdoor, mixed distributions   * Configurations: different beam patterns |
| CMCC | Different beam measurement configurations (e.g., different number of measured beam pairs in set B) | We can set 16, 32，48, 64 as the number of measured beam pairs in set B.  At least applicable to BM-Case1. |
| Ericsson | We encourage companies to provide a wide area of generalization evaluations.  Examples could include:  -different antenna config  -different propagation (via spatial consistency)  -different cell/site  -different UE mobility  …. | We encourage companies to provide results for many generalization alternatives. However, it would be beneficial to outline a set of prioritized generalization metrics. |
| Samsung | For gNb side training/inference:   1. Different UE speed, UE distribution, number of Rx when applicable.   For UE side training/inference:   1. Difference channel model 2. Different gNB Tx beam parameters, # of Tx beam/ pattern of Set A/Set B   Different UE speed, UE distribution | Training/inference location needs to be considered. |
| Lenovo | Some of the scenarios/configurations/parameters that need to be considered for testing the generalizability are as follows:   * Different number of Tx beams * Different number of Rx beams * Different beam widths * Beam angles * Overlapping/non-overlapping Tx beams * Different UE speeds   Different channel model configurations such as LoS/NLoS, indoor/outdoor etc., covering both the cases of channels with single dominant path and channels with multipath transmission. |  |
| Qualcomm | Inter-site (heterogeneous): train AI/ML model on a first set of deployment type(s) and test it on a second (unseen) deployment type.  Inter-site (homogeneous): train on a first set of site(s) of a given deployment type and test it on a second (unseen) site of that same deployment type.  Intra-site: train AI/ML model for a given site and test it on unseen variations within that same site.  Across configurations: train AI/ML model on a first set of configuration(s) and test on a second configuration | Inter-site (heterogeneous): train AI/ML model on a first set of deployment type(s) and test it on a second (unseen) deployment type.   * 1. Example: train for UMi deployment and test on UMa deployment   Inter-site (homogeneous): train on a first set of site(s) of a given deployment type and test it on a second (unseen) site of that same deployment type.   * 1. Example: train and test on two different realizations of UMi 🡪 This can be realized by considering different random seeds within the same deployment which would represent a different ‘environment’ within that deployment.   Intra-site: train AI/ML model for a given site and test it on unseen variations within that same site. Let us consider two sub-categories:   * 1. Environment variations due to moving objects, time of day, foliage variation over seasons, etc.: Given the agreements regarding utilizing channel models based on 38.901 for evaluations, the existing channel models do not support this systematic intra-site dataset generation.   2. Basic generalization scenarios: Scenarios identifying minimum conditions/requirements for an AI/ML model to work well in practice      1. Examples for these scenarios are training and testing on different sets of UE speeds/orientations/trajectories   Across configurations: train AI/ML model on a first set of configuration(s) and test on a second configuration   * 1. Example: train and test for different UE or gNB codebooks or different gNB array down-tilt angles.   For downlink TX beam prediction, we can train assuming first UE codebook and test assuming second UE codebook, and then, we can try to see how well the AI/ML model generalizes across these two scenarios. This is of practical significance particularly for gNB-side beam prediction to see how well an AI/ML model that has been trained for a certain UE ‘type’ generalizes to another UE ‘type’. Now, let us consider UE-side AI/ML models for beam prediction. We can train an AI/ML model at the UE side assuming first gNB codebook and test the AI/ML model assuming second gNB codebook (potentially with different beam shapes) and see how the AI/ML model performs across these scenarios. This is also of practical significance, as we can analyse how the UE-side AI/ML model that has been trained in a given cell generalizes to a secondary cell within the same site in which the secondary gNB uses a different codebook, potentially with different beam shapes. |
| HW/HiSI | For BM-Case 1, the scenarios/configurations for performing the inference for the AI/ML model include at least:   * Various channel types, including UMa, UMi, InH * Various indoor/outdoor ratios, including 10:0, 8:2, 5:5, 2:8, 0:10 * Various numbers of beams in Set A/Set B   For BM-case 2, the scenarios/configurations for performing the inference for the AI/ML model include at least:   * Various channel types, including UMa, UMi, InH * Various indoor/outdoor ratios, including 10:0, 8:2, 5:5, 2:8, 0:10 * Various UE speeds (e.g., 30km/h, 60km/h, 90km/h) * Various types of trajectories   Various numbers of beams in Set A/Set B | In our paper, we already provide sufficient discussion on how to evaluate generalization performance and the location can be:  To verify the generalization of AI/ML models on AI/ML-based beam management in both BM-Case1 and BM-Case2, the following cases to construct the training dataset and testing dataset should be considered:   * Case 1: The AI/ML model is trained based on training dataset from one Scenario#A/Configuration#A, and then the AI/ML model is tested on dataset from the same Scenario#A/Configuration#A * Case 2: The AI/ML model is trained based on training dataset from one Scenario#A/Configuration#A, and then the AI/ML model is tested on dataset from a different Scenario#B/Configuration#B * Case 3: The AI/ML model is trained based on training dataset constructed by mixing datasets from multiple scenarios including Scenario#A/Configuration#A and Scenario#B/Configuration#B, and then the AI/ML model is tested on dataset from a single Scenario#A/Configuration#A or Scenario#B/Configuration#B from the multiple scenarios * Case 4: The AI/ML model is trained based on training dataset from one Scenario#A/Configuration#A, and then the AI/ML model is fine-tuned based on the fine-tuning dataset from a different Scenario#B/Configuration#B. After that, the AI/ML model is tested on dataset subject to the same Scenario#B/Configuration#B as the fine-tuning dataset |

## 1.4 AI/ML model related aspects

The following proposals are collected from contributions on AI/ML model:

* Vivo [3]: At least AI model inputs/outputs and training/validation dataset should be reported per sub-use case by companies. Other parameters, such as NN architecture type, loss function, and data post/pre-processing method, are encouraged to be reported.
* ZTE [4]: For cross-checking and reproducibility purposes, a high-level description of the adopted AI/ML model such as NN architecture type, model input/output, and training methodology should be disclosed by each company
* Samsung [17]: At least AI/ML model structure, input /output of AI/ML model shall be reported by companies.
* Ericsson [20]: To help enable reproducibility, companies are encouraged to report relevant information about the AI/ML model architecture, data pre- and post-processing, loss functions, and training procedures using an academic style paper and/or pseudocode.

Moreover, the summary of AI models structure used by companies for two use cases can be found as:

Spatial domain prediction:

* CNN: Futurewei [1], NVDIA [12], Intel [12], Nokia [21], CMCC [19] (CNN+FC)
* DNN: OPPO [9], Intel [12], Nokia [21], CEWiT [23], CATT [10] (DNN based and ResNet based)
* LSTM +BN+FC: Samsung [17]:
* FNN: DoCoMo [26]
* Full connection: Vivo [3], ZTE [4], Fujitsu [5], China Telecom [15]

Temporal prediction:

* RNN: NVDIA [12]
* LSTM:
  + ZTE [4]: LSTM, FC layer
  + OPPO [9]: LSTM (Long-Short Term Memory) + DNN models
  + Samsung [17]: LSTM +BN+FC
  + Nokia [21]: LSTM-based, Conv2D
  + Qualcomm [24]: LSTM
  + DoCoMo [26]: LSTM

Based on the summary, the following proposals can be discussed.

#### FL1 (Low) Question 1-4-1a

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

* **Description of AI/ML model, e.g, NN architecture type, shall be reported by companies for evaluation of AI/ML in BM.**

|  |  |
| --- | --- |
| Supporting companies | MediaTek, CAICT, vivo, Futurewei (with comments), CATT, CMCC, FUJITSU, ZTE,DCM, Ericsson,Samsung, Lenovo, Qualcomm (with comments), HW/HiSi |
| Objecting companies |  |

**Question 1-7a:**

Please provide your view on **Proposal 1-4-1a**

|  |  |
| --- | --- |
| Company | Comments |
| FL1 | Please also provide your views if any other aspects shall be reported by companies. |
| Futurewei | Companies are encouraged to provide their NN model type, but no need to mandate it. |
| Lenovo | It is desirable to provide enough information to enable reproducibility of the results by other companies. |
| Qualcomm | We support reporting any information about the AI/ML model as long as it is left up to companies (arbitrary). |

## 1.5 Other consideration for evaluations

# KPIs on AI/ML in beam management

Two types of KPIs were proposed by companies for evaluation. One is performance related KPIs, the other is AI/ML model capability related KPIs including size of the models, computation complexity, generalization, etc.

## 2.1 Beam prediction accuracy related KPIs

In RAN 1 #109e, several beam prediction accuracy related KPIs were identified. Based the contributions from companies, the KPIs used/proposed were summarized as below:

* + Beam prediction accuracy related KPIs, may include the following options:

|  |  |  |  |
| --- | --- | --- | --- |
| **Beam prediction accuracy related KPIs** | | | **Companies support or choose this KPIs** |
| Agreed | * Average L1-RSRP difference of Top-1 predicted beam   + Definition: The difference between the ideal L1-RSRP of Top-1 predicted beam and the ideal L1-RSRP of the Top-1 genie-aided beam | | FUTUREWEI, vivo, ZTE, Fujitsu, Interdigital, CATT, CAICT, China Telecom, Samsung, Nokia, Qualcomm |
| * Beam prediction accuracy (%) for Top-1 and/or Top-K beams, FFS the definition: | * Option 1: The beam prediction accuracy (%) is the percentage of “the Top-1 predicted beam is one of the Top-K genie-aided beams” | Intel, CAICT, China Telecom, Samsung, Qualcomm |
| * Option 2: The beam prediction accuracy (%) is the percentage of “the Top-1 genie-aided beam is one of the Top-K predicted beams” | FUTUREWEI, Huawei, vivo (1 and K=4), ZTE, Fujitsu, OPPO, CATT, Intel, Xiaomi, Samsung, CMCC, Nokia, CEWiT, Qualcomm |
| * CDF of L1-RSRP difference for Top-1 predicted beam | | FUTUREWEI, Fujitsu, CATT, CAICT, China Telecom, Apple |
| * Beam prediction accuracy (%) with 1dB margin for Top-1 beam   + Definition: The beam prediction accuracy (%) with 1dB margin is the percentage of the Top-1 predicted beam “whose ideal L1-RSRP is within 1dB of the ideal L1-RSRP of the Top-1 genie-aided beam” | | Vivo, ZTE, Interdigital, Intel, Samsung, Nokia, Qualcomm |
| New | * Average L1-RSRP difference of Top-K predicted beam   + Definition: the difference between the highest ideal L1-RSRP of the Top-K predicted beam and the ideal L1-RSRP of the Top-1 genie-aided beam. | | FUTUREWEI |
| * The difference between the predicted L1-RSRP of Top-1 predicted beam and the ideal L1-RSRP of the Top-1 genie-aided beam | | OPPO |

Moreover, some observations/proposals were made in the contributions:

* Futurewei [1]
  + Observation 4: When evaluating AI/ML model performance, using “Average L1-RSRP difference of Top-1 (or Top-K) predicted beam” alone may not directly indicate the performance unless the average L1-RSRP difference between the ideal L1-RSRP of the Top-1 genie-aided beam and the ideal L1-RSRP of the Top-K genie-aided beams in the (testing) dataset is known.
* Huawei [2]
  + Proposal : Since the prediction accuracy obtained from the AI/ML increases significantly with a larger K and then clearly outperforms the legacy baseline, adopt Top-K, K>1 (e.g. K=3,5) for evaluation of spatial beam prediction accuracy.
  + Proposal : For temporal beam prediction evaluation, results for Top-K, K>1 should be presented in addition to Top-1 results.
    - The Top-1 predicted beam can be derived as the eventual result after the second round sweeping based on the AI/ML inferred Top-K beams.
* Vivo [3]
  + Observation 1: If only measured RSRP is input into the AI model, large performance loss appears in the case that training and inference use different sets for measurement.
  + Observation 2: Fixed beam subset in Set B can have good performance in ideal scenarios but it lacks flexibility. Issues like blockage and inter-cell interference can bring negative impact on the performance of fixed subset.
  + Observation 3: For random subset selection, i.e., training and inference use different beam subsets in Set B, to include Tx/Rx beam ID or angle into the AI model is helpful to reduce performance loss.
  + Observation 4: To restrict the selection of random subset from the best X beam subsets can improve the performance of BM Case 1 prediction. Such semi-random selection with Tx/Rx beam angle information as input barely suffers performance loss compared with the best beam subset.
* MediaTek [22]:
  + Proposal 1: For AI/ML-based beam prediction evaluation, adopt the top-k beam prediction accuracy as the intermediate performance, and the RSRP gain comparing to the baseline and upper bound method as the eventual performance metric.

#### FL1 (Low) Question 2-1-1a

**Question 2-1-1a: whether the following new L1-RSRP related KPIs can be defined:**

* KPI #1: Average L1-RSRP difference of Top-K predicted beam
  + Definition: the difference between the highest ideal L1-RSRP of the Top-K predicted beam and the ideal L1-RSRP of the Top-1 genie-aided beam.
* KPI #2: The difference between the predicted L1-RSRP of Top-1 predicted beam and the ideal L1-RSRP of the Top-1 genie-aided beam

|  |  |  |  |
| --- | --- | --- | --- |
| **Company** | **KPI#1** | **KPI#2** | **Comments** |
| FL1 | Y or N | Y or N | Comments, if any, including whether down selection is needed for L1-RSRP related KPIs? |
| MediaTek | N |  | Current agreed KPIs are enough (In the current agreement, we already have a variety of KPIs that are enough to evaluate the prediction accuracy for both BM-Case1 and BM-Case2) |
| Xiaomi |  | Y | It is necessary to evaluate the accuracy of L1-RSRP prediction. |
| vivo | N | N | We think more discussion is needed on why we need to add these two definitions on top of the options we already have in last meeting. |
| Futurewei | Y | Y | We are ok to define them while companies can choose what KPIs they want to use/report together with their results. |
| CATT | N | N | The current KPIs are enough. |
| Fujitsu | N | N | * For KPI#1, our understanding is the highest ideal L1-RSRP of the Top-K predicted beam equals to ideal L1-RSRP of Top-1 predicted beam. In this case, there is not difference between KPI#1 and average L1-RSRP difference of Top-1 predicted beam. |
| ZTE | Y | Y | KPI#1 can be defined. If a second stage beam sweeping over the Top-K predicted beams is conducted, this KPI can reflect the gap between the final beam and the ideal genie-aided beam.  KPI#2 can be defined. The predicted L1-RSRP can be considered for KPI definition. Considering that Top-K beams are predicted by the AI model. If a second stage UE-specific beam sweeping over these Top-K beams is conducted, the resulting RS overhead might be huge, which significantly diminishes the advantage of the AI approach. Instead, gNB may directly select a final beam for data transmission within the Top-K beams. In this case, the predicted L1-RSRP can provide a good reference for the beam selection. |
| NTT DOCOMO | Y | Y | We are fine with defining new KPIs as optional KPIs, even though we think the existing intermediate KPI is enough, |
| Ericsson | N | N | Current KPIs are enough |
| Samsung | N | N | Current list of KPIs is sufficient. |
| Lenovo | Y | N | Rather than “predicted” L1-RSRP, we prefer to find out the difference between the “ideal” L1-RSRP values. |
| Qualcomm | Y | N | KPI #1 is consistent with the prior agreement (109e) on L1-RSRP difference of Top-1 predicted beam. |
| HW/HiSi | Y | N | For KPI#2, with top-1, is it meant the AI model is inferring one beam, or does it mean the final beam after 2nd stage beam sweep? When we answered “N”. we assumed the former |

## 2.2 System performance related KPIs

### 2.2.1 User throughput (on hold)

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

* Interdigital [6]: Proposal 2: Support system performance related KPIs as mandatory KPIs.
  + Support Avg. and 5% UE tput for system performance KPIs.
* Samsung [17]: Proposal 8: Shannon capacity-based simplified model for UPT can be further considered as additional system performance related KPI.
* Nokia [21]: Proposal 9: For BM-Case1, RAN1 further investigate RL-based beam prediction to see the possible gains on system throughput.
* Qualcomm [24]Proposal 7: At least for spatial domain beam prediction, consider spectral efficiency CDF for SLS evaluations as a KPI.

### 2.2.1 RS overhead

There were several proposals/discussions related to RS overhead:

* Huawei/HiSi [2]: Proposal 13: For the overhead calculation of temporal domain beam prediction, the observation window and prediction window should be considered jointly, and the overhead reduction compared to a reference should be regarded as averaged combined reduction.
* Vivo [3]: The metric of beam sweeping overhead reduction is calculated as 1-N/M where N is the number of beams required for measurement in both non-AI algorithm and AI algorithm, and M can be the total number of all possible beams to be predicted.
* ZTE [4]:RS overhead reduction can be considered as a basic KPI for evaluation and should be further studied with factors to be considered including: the number of UEs, the beam pattern, and the refined beam sweeping procedure.
  + RS overhead reduction can be defined as 1-N/M with N being the number of beams for measurement and M being the number of beams for prediction
* Fujitsu [5]: Proposal 3: For spatial-domain DL beam prediction in mTPRs scenario, the following intermediate KPIs are suggested for each TRP respectively.
  + RS overhead reduction
* Interdigital [6]: Proposal 4: Reporting overhead and latency aspects should be considered in evaluation of system performance not as independent KPIs.
* Lenovo [11]: Proposal 3: Consider beam prediction accuracy, Latency reduction and RS overhead reduction as the key KPIs in evaluating any AI/ML model for beam management.
* Ericsson [20]: Define a RS measurement reduction KPI, e.g. 1-N/M where N is the number of beams measured by a UE, and M is the total number of beams.
* Nokia [21]: RS overhead reduction at least for spatial-domain beam prediction at least for Top-1 beam

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

* DoCoMo [26]: Proposal 1: Discuss whether additional RS overhead should be considered, when the top-1/K predicted beam(s) are not included in beams measured for the beam prediction.
  + For example, the following equation can be considered as KPI for RS overhead reduction.

Based on the discussion, the following proposals can be discussed:

#### FL1 (High) Question 2-2-1a

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

* For RS overhead reduction, further study the following options:
  + Option 1:
    - where N is the number of beams (with reference signal (SSB and/or CSI-RS)) required for measurement
    - where ~~(FFS)~~ M is the total number of beams to be predicted (in Set A)
  + Option 2:
    - Where is the number of beams required for measurement during time slot
    - where M is the total number of beams to be predicted (in Set A)
  + Option 3:
  + Other Options are not precluded

|  |  |
| --- | --- |
| Option 1 Supporting companies | OPPO, MediaTek, CAICT, vivo, Futurewei, CATT, FUJITSU, ZTE, Ericsson, Samsung, Qualcomm |
| Option 2 Supporting companies | Vivo,Samsung, Lenovo |
| Option 3 Supporting companies | DCM |

**Please provide your view Proposal 2-2-1a, including the definition of M**

|  |  |
| --- | --- |
| Company | Comments |
| FL1 | The definition of M needs to be clarified, e.g., whether this is number of Tx beams, or number of Tx/Rx beam pair? Or this could subject to the outputs/inputs? |
| OPPO | From our reading, the beams mentioned in Proposal 2-2-1a seem quite generic, i.e. possibly referencing to Tx beam, Rx beam or Tx-Rx beam pair. In this proposal, the intention is to address how to calculate overhead reduction and it can be subject to input/output of AI/ML model. |
| MediaTek | Option1 is simple and enough |
| Xiaomi | First, we would like to clarify that this proposal is for BM Case 1 only? For BM case 1, we prefer Option 1.  While for BM case 2, we think the RS overhead should be calculated differently from BM case 1. For BM –case 2, if set B=set A, the RS overhead can be 100% or with K is the number of latest measurement instance and F is the number of future time instance, when the periodicity is same for latest measurement instance and future time instance. If the periodicity is different for latest measurement instance and future time instance, different options may be considered. |
| vivo | We think Option 1 can be sufficient for BM Case 1, and Option 2 can provide more information in BM Case 2. |
| CMCC | In our view, the predicted Top-K beam pairs will be measured by UE for best beam pair selection and obtain the actual QCL relationship . Hence, the RS overhead reduction can be |
| ZTE | In Option 1, whether N includes the RS overhead for measuring the Top-K beams needs to be clarified. If Top-K beams are predicted by the AI model, how to obtain a final beam for data transmission can be discussed and may significantly impact the RS overhead calculation. For example, gNB can randomly select a beam within the Top-K beams or conduct a second stage UE-specific beam sweeping. The associated RS overhead would be different. |
| NTT DOCOMO | Could you remove the sigma function on the nominator in Option 3?  For Option 1/2, the number of beams required for measurements is unclear. If it means that the number of beams required before the PDSCH/PDCCH reception, we are fine with Option1/2. If it means the number of beams required for identifying the top1-predicted beams, we do not support Option 1/2 at this stage. Our concern is that the actual L1-RSRP measurements (non-predicted L1-RSRP) might be necessary before the PDSCH/PDCCH reception. In that case, just identifying the best beam by beam prediction is not enough for UE reception with the corresponding beam. To avoid the overestimation of beam prediction performance, we would like to discuss if the actual L1-RSRP measurements is necessary or not at first. |
| Samsung | We think both option1 and 2 can be considered.  One more issue needs to be clarified, whether M is beam or beam pair. |
| Lenovo | Prefer Option 2, as it supports the more generic case where the number of beams to be measured varies from time to time. Value of M depends on the type of input to the AI/ML model (discussed in one of previous questions). |
| HW/HiSi | Not support.  Option 1 does not give a fair comparison in overhead. Assume that Set B consists of 16 beams and set A of 32, 64 or 256.   * For Set A = 32: the overhead is 1-16/32=0.5 * For Set A = 64: the overhead is 1-16/64=0.75 * For Set A = 256: the overhead is 1-16/256=0.9375   But in fact the cost is the same for all sizes of Set A, since always 16 beams will be swept regardless the size of Set A. Similar is valid for Option2 and also for Option 3.  In our understanding the size of Set A does not determine the overhead or better the cost What is important to measure is how many beams are included in Set B, because these are the number of beams that have to be swept in P1 and then the number of K in Top-K which are swept in P2.  We propose an Option 4, to not calculate a relative overhead, since it is difficult to define “relative to what”, and instead to count the number of beams that have to be swept:  Option 4: RS cost = N for K=1, N+K for K >1, where N is the number of beams in set B |

### 2.2.3 Other KPIs

Latency

There were several proposals/discussions related to latency overhead:

* Lenovo [11]: Proposal 3: Consider beam prediction accuracy, Latency reduction and RS overhead reduction as the key KPIs in evaluating any AI/ML model for beam management.
  + Definition:
    - 1– ([Total transmission time of *N* beams] / [Total transmission time of *M* beams])
      * where *N* is the number of beams (with reference signal (SSB and/or CSI-RS)) in the input beam set required for measurement and *M* is the total number of beams
* CMCC [19]: Proposal 2: If latency refers to the latency of transmission of measured beam pairs, the definition of latency reduction is 1 – [Total transmission time of N beams] / [Total transmission time of M beams], where the transmission time of beams starts from the earliest symbol of the measured RS and ends at the latest symbol of the RS within one period of RS transmission.

UCI report

* Vivo [3]: UCI reporting overhead reduction, including the number of UCI report and UCI payload size, should be considered as basic KPI.
* DoCoMo [26]: Encourage companies to provide the reporting overhead, e.g., the number of uplink transmissions for CSI reports and UCI overhead, for the temporal beam prediction with NW side model.

RRC signaling

* Vivo [3]: RRC singling overhead can be considered as optional KPI if huge amount of data, such as training data, assistant information, and AI model data, is exchanged via RAN air interference.

#### FL1 (Low) Question 2-2-2a

**Question 2-2-2a: whether the following KPIs needs to be defined, and how:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Company** | **Latency** | **UCI report** | **RRC signaling** | **Comments** |
| FL1 | Y or N | Y or N | Y or N | If yes, please provide preferred definition, if any |
| MediaTek | Y | Y | N | Latency: we think latency is important and should be included as KPI for AI BM, the definition can be FFS  UCI report: the definition of UCI report KPI can be based on the overhead for reporting the measured RSRP for model training. |
| vivo | Y | Y | Y | UCI: include number of UCI reports and UCI payload size  RRC signaling: RRC signaling size |
| NTT DOCOMO |  | Y |  | We suggest UCI payloads and the number of UCI reports as KPIs.  UCI overhead evaluation should be considered for beam prediction with NW side model. The beam prediction at NW side might require UCI including a large number of L1-RSRP and assistance information for beam prediction. For example, if the Rx beam information is included in UCI reports for beam pair prediction, it could enlarge the UCI reports, which reduces the resource efficiency. To analyze the performance of beam prediction fairly, we prefer to add UCI payload as KPI.  Also, the number of uplink transmissions due to UCI reports is an important factor in NW operation. As temporal beam prediction has a potential to reduce the frequency of beam measurement reports by compensating with the beam prediction, we should evaluate the performance as well. |
| Ericsson | Y | Y | Y | Definitions are FFS |
| Samsung | FFS | Y | N | Some further study on those aspects is needed.  We think the it may impact on the assumption of beam management procedure, e.g., whether UE needs sweeping to obtain the measurement of a Tx beam. Some further clarification is needed  For UCI report, it has two level, one is the number of reports in a given time, the other is the bits in each UCI report. We think this may impact on the BM performance.  For RRC signaling, is this for model transfer /data collection or something else? This can be discussed together with the necessary for“online data collection”and “model transfer” for BM. |
| Lenovo | Y | Y | Y | For latency, please consider our proposal [11]. For UCI report and RRC signaling, we think Vivo’s definitions [3] works well. |
| HW/HiSi | N | N | N | Latency can maybe be considered later optionally, at least for temporal BM.  UCI and RRC signalling do not need to be included in the SI phase. This can be left for work item. |

## 2.3 Model size and computational complexity

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

* Vivo [3]: Model complexity, computational complexity, and power consumption are all relevant and can be estimated using number of floating-point operations (FLOPs) as well as AI model size in terms of number of parameters or bytes.
* Fujitsu [4]: Proposal 2: For the AI/ML model complexity, the following KPIs are selected
  + The number and data type of parameters on AI/ML model
  + The number of float point of operations for AI/ML model inference
* Lenovo [11]: Complexity of the proposed AI/ML model should be evaluated for every phase in the model lifecycle, namely, training, inference and update.
* NVIDIA [12]:
  + Proposal 3: For evaluation of AI/ML based beam management, the computational complexity can be reported via the metric of floating point operations (FLOPs).
  + Proposal 4: For evaluation of AI/ML based beam management, the model complexity may be measured by memory storage in terms of AI/ML model size and number of AI/ML parameters.
  + Observation 1: Increasing hardware performance can support successively more complex AI/ML models. For example, GPU inference performance has improved by 317x in 8 years (2012-2020), more than doubling each year.
  + Proposal 5: AI/ML model complexity and computational complexity should not be regarded as a roadblock to the adoption of AI/ML based algorithms for beam management enhancements.
* Samsung [17]:
  + Proposal # 9: For the use case of AI/ML based beam management, at least the following capability-related KPI shall be considered:
    - Size of AI/ML model;
    - Complexity of training and inference of AI/ML operation.
* LGE [18]: Proposal 3. For complexity KPI for AI/ML in BM, at least FLOPs, memory usage and complexity of pre-processing can be considered.
* Ericsson [20]: When presenting results for AI/ML models, the proponent should report a model size (e.g., number of parameters) and an estimate of the number of floating-point operations (FLOPs) for inference.
* MediaTek [22]:
  + Proposal 2: For AI/ML-based beam prediction evaluation, adopt the FLOPs and/or MACs as the time complexity, and the number of parameters and actual model size as the space complexity.
  + Proposal 5: Study and evaluate other forms of ML models with respect to their prediction performance and computation complexity for spatial beam prediction.

The following had been reported by companies:

Futurewei [1]:

|  |  |  |
| --- | --- | --- |
| Total number of beams | Number of NN parameters | FLOPs |
| 256 | 2,491,968 | 2,605,568 |

ZTE[4]:

|  |  |
| --- | --- |
| Spatial domain  Model size | 363.264K (for Case 1&3)  1.646M (for Case 2&4) |

Fujitus[5]

|  |  |  |  |
| --- | --- | --- | --- |
|  | Number of float point of operations | Number of parameters | Memory size |
| FC (figure 2) | 213.4K | 214K | 857KB |
| FC + CNN (figure 2-A) | 14.6M | 270K | 1.1MB |

CATT[10]

|  |  |
| --- | --- |
| Model Size | Model 1(DNN): 430k |
| Model 2(ResNet): 280k |

Samsung [17]

* Observation #1: The Params of the AI/ML model used in the simulation are about Params whose memory occupation is about 15MB.
* Observation # 2: FLOPs of the AI/ML model is about .

CMCC [19]

|  |  |  |
| --- | --- | --- |
|  | FLOPs(×106) | Trainable Parameters(×106) |
| 12\*4 | 3.74 | 1.87 |
| 8\*4 | 3.21 | 1.61 |
| 4\*4 | 2.69 | 1.35 |

Ericsson [20]

|  |  |
| --- | --- |
| Model complexity KPIs | * Number of parameters:   + 4x8 gNB array with 8 SSB beams in Set B:     - NN A: ~1200     - NN B: ~9000   + 8x16 gNB array with 21 SSB beams in Set B:     - NN A: ~32000 * Floating point operations for inference   + 4x8 gNB array with 8 SSB beams in Set B:     - NN A: ~2500 FLOPs     - NN B: ~17000 FLOPs   + 8x16 gNB array with 21 SSB beams in Set B:     - NN A: ~64000 FLOPs |

CEWiT [23]

|  |  |
| --- | --- |
| Model Complexity | 68.18 k FLOPS |

DoCoMo [26]

|  |  |  |
| --- | --- | --- |
| FLOPs | Memory size | Param. size |
| 270K | 700MB | 271K |

#### FL1 (High) Question 2-3-1a

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

* **For the evaluation of the AI/ML in BM, the following AI/ML model related KPIs can be considered and reported by companies.**
  + **Floating point operations (FLOPs) for inference**
  + **Size of AI/ML model, FFS:**
    - **Number of parameters**
    - **Bytes**

|  |  |
| --- | --- |
| Supporting companies | OPPO, MediaTek, CAICT, vivo, Futurewei, CATT, CMCC, Fujitsu, ZTE,DCM, Ericsson, Samsung, Lenovo, Qualcomm (with comments), HW/HiSi |
| Objecting companies |  |

**Please provide your view Proposal 2-3-1a, if any.**

|  |  |
| --- | --- |
| Company | Comments |
| Qualcomm | The overall complexity of an AI/ML model heavily depend on implementation details and metrics such as FLOPs, number of parameters etc. are only (very) loosely representative of the actual complexity of AI/ML models for implementation. Also we believe this aspect should be voluntarily reported by companies. |

## 2.4 Baseline performance (on hold)

Some companies provided some analysis on baseline performance for benchmark.

**Spatial domain beam prediction**

The following proposals/discussions were provided in the contributions on baseline performance for spatial domain beam prediction:

* Vivo [3]: Support both option 1 and option 2 as baseline performance in spatial domain beam prediction and temporal domain beam prediction, and subset selection method in option 2 should be reported.
* ZTE [4]:
  + Proposal 1: For spatial domain beam prediction, the case that the measured RSRPs of partial beam pairs are used as the AI input and the predicted RSRPs of all beam pairs are used as the AI output can be adopted as a baseline for comparison.
  + Proposal 5: For temporal beam prediction, the case that the measured RSRPs of all beam pairs of past time instances are used as the AI input and the predicted RSRPs of all beam pairs of future time instances are used as the AI output can be adopted as a baseline for comparison.
* Interdigital [6]:
  + Observation 1: Legacy beam management with Rel-17 without AI/ML algorithms is not an appropriate baseline as implementation-based AI/ML operation is available for UE and gNB implementations.
  + Proposal 1: ‘No collaboration framework: AI/ML algorithms purely implementation based and not requiring air-interface changes’ could be an appropriate baseline to accurately evaluate the benefits of AI/ML with specification enhancements.
* Google [7]: Proposal 4: For spatial-domain beam prediction, the baseline performance should be the performance from the beam selected from set B beams.
* OPPO [9]: Proposal 3: For spatial domain beam prediction, select the best beam within Set A via exhaustive beam sweeping (Option 1) as baseline.
* Lenovo [11]: Select “Option 1: Select the best beam within Set A of beams based on the measurement of all RS resources or all possible beams of beam Set A (exhaustive beam sweeping)" as the baseline for spatial beam prediction.
* Intel [13]: For baseline performance evaluation, Option 2 should correspond to hierarchical beam search where, based on sub-use case being evaluated, set B may be a subset of set A or set B can contain both wide and correlated narrow beams.
  + The hierarchical beam search consists of two steps. First, x beams are uniformly sampled from the 64 input beams, and the beam corresponding to the largest RSRP is selected. Second, y beams around the selected beam in step 1 is selected, and the beam corresponding to the largest RSRP is selected as the final beam search output.
  + The hierarchical beam search consists of two steps. First, 16 beams are uniformly sampled from the 16 input wide beams, and the beam corresponding to the largest RSRP is selected. Second, 4 narrow beams corresponding to the selected wide beam in step 1 is selected, and the beam corresponding to the largest RSRP is selected as the final beam search output.
* Xiaomi [14]: Here we take Option 2 as the baseline and consider Alt.1 that set B is a subset of set A.
* CAICT [15]: Proposal 2: For spatial-domain beam prediction, select the best beam within Set A based on the measurement of Set B as baseline.
* Samsung [17]: In this scenario, we select the best beam in Set B of beam as the predicted Top 1 beam as baseline performance, by assuming the same measurement/resource are used by UE. Therefore, the baseline performance depends on whether the best beams fall into the pre-defined beams in Set B.
* LGE [18]: For option 2, assumptions on Set A and Set B configuration and how conventional scheme to obtain performance KPIs are reported by each company.
* CEWiT [23]
  + Proposal 1: Baseline performance evaluation for comparing the performance of the AI/ML model should be based on exhaustive beam sweeping, i.e., support option 1 as the baseline performance scheme.
  + Proposal 2: Support use of CSI-RS for beam measurements in the baseline scheme.
* Qualcomm [24]
  + Baseline-1 (linear interpolation-based approach): For the non-measured beams in Set A, we estimate their RSRPs by performing a 2D linear interpolation (and extrapolation, when applicable) based on the measured RSRPs and the pointing angle of the beams. We select the top K predicted beams based on the interpolated RSRPs.
  + Baseline-2 (empirical PMF-based approach): We derive an empirical probability distribution of the best beam ID in Set A, given the best beam ID in Set B, based on the training dataset. For testing, we select the top K predicted beams based on the derived empirical distribution.

**Temporal beam prediction**

* Google [7]: Proposal 5: For time-domain beam prediction, the baseline performance should be the performance without beam change for T2, i.e. the beam used prior to T2 is applied for T2.
* OPPO [9]: Proposal 4: For temporal domain beam prediction, select the best beam for T2 within Set A via exhaustive beam sweeping (Option 1a) as baseline.
* Xiaomi [14]: Here we take Option 1a as the baseline and consider Alt.3 that set A and set B are the same.
* CAICT [15]: Proposal 3: For temporal beam prediction, Select the best beam for T2 within Set A of beams based on the measurements of all the RS resources or all possible beams from Set A of beams at the time instants within T2 as baseline.

# Evaluation results for AI/ML in beam management

## 3.1 Evaluation results for spatial domain prediction (on hold)

The following observations were provided in contributions:

Futurewei [1]:

* Observation 3: We observe that AI/ML based spatial beam prediction achieved better performance compared to sparse beam sweeping approach.

Huawei [2]:

* Observation 1: For spatial domain beam prediction, AI/ML-based schemes under the 64-DFT codebook outperform the legacy approach in most of the cases in terms of beam selection accuracy, e.g.,:
  + AI/ML-based Top-5 prediction reaches almost the upper performance bound with a prediction accuracy of 94.95% but with an overhead reduction of 67.17%. On the other hand, for the same overhead reduction, the established legacy Baseline approach can only achieve a prediction accuracy of 55.3%
  + With AI/ML-based Top-3 prediction, the overhead compared to the legacy Baseline approach can be further reduced by another 8%, while the prediction still is much higher (89.2% as opposed to 55.3%)
* Observation 2: For spatial domain beam prediction, AI/ML-based schemes under the 64-DFT codebook outperform the legacy approach in most of the cases in terms in terms of average L1-RSRP difference, e.g.,:
  + For AI/ML-based Top-5 prediction, the L1-RSRP difference compared to genie aided beam prediction in Exhaustive 64 is as low as 0.03 dB, with an overhead reduction of 67.17%. On the other hand, for the same overhead reduction, the established legacy Baseline approach can only achieve an average L1-RSRP difference of 1.02dB
  + With AI/ML-based Top-3 prediction, the overhead compared to the legacy Baseline approach can be further reduced by another 8%, while the average L1-RSRP difference is still is much smaller (0.08dB as opposed to 1.02dB)
* Observation 4: The AI/ML-based beam prediction based on the Set A with 256 beams (Type-2) provides a considerable gain over the legacy upper bound Exhaustive 64 (Type-1) in achieved L1-RSRP and cost with a small fraction of the overhead.

ZTE [4]:

* Observation 3: The spatial domain beam prediction with the fixed beam pattern achieves a sufficiently high performance with only 25% beam overhead being used.
* Proposal 4: The AI/ML model can be utilized for spatial domain beam prediction, which can greatly reduce the RS overhead for measurement while maintain a high beam prediction accuracy.

Fujitsu [5]:

* Observation 2: For sTRP scenario,
  + Comparing to the exhaustive beam sweeping, the performance of AI/ML model is
    - The RS overhead is reduced 75%.
    - The beam prediction accuracy of Top-2 is about 80%.
    - The average L1-RSRP difference is less than 2 dB.
    - The probability is about 70% with the estimated L1-RSRP difference for Top-1 predicted beam less than 2dB.
  + For the complexity of AI/ML model,
    - The AI/ML model has total 214K parameters with single-float data type.
    - The computational complexity of AI/ML model is about 213.4K float point of operations.

Interdigital [6]:

* Observation 4: AI aided beam selection achieves more than 95% selection accuracy when error margin is larger than 0.5 dB by consuming 50%/33% of the measurement overhead for the exhaustive measurement.
* Observation 5: AIML-based RSRP estimation always outperforms the baseline especially when less RSRP measurements are available as it achieves a higher selection accuracy by 35% when error margin is 0.5 dB.

Rakuten Mobile Inc. [8]:

* Observation 1: The probability of one of the K beams being the best beam is more than 95% for K = 4.

OPPO [9]

* Observation 1: Spatial domain beam prediction can yield beam prediction accuracy (at least 80%) while overhead/latency reduction rate is 75%.
* Observation 2: The system level metric, i.e. spectrum efficiency or throughput, is not sensitive to the L1-RSRP difference introduced by spatial domain beam prediction.
* Observation 3: For 80% of the incorrect spatial domain beam prediction cases, the L1-RSRP difference can be kept within 2dB.
* Observation 4: When beam prediction accuracy is high and L1-RSRP difference is small, the performance evaluation on system-level output, i.e. spectrum efficiency or throughput may only shed light on non-essential insight.
* Observation 8: For InH scenario, AI/ML beam prediction in spatial domain can yield relatively high correct prediction rate (around 80%) while using only a small portion (1/8) of DL measurement.
* Observation 9: For InH scenario, when AI/ML beam prediction is correct, the L1-RSRP prediction error is acceptably small (around 1dB); otherwise (AI/ML beam prediction incorrect), the L1-RSRP prediction error would increase slightly (up to 2.5dB).

CATT [10]:

* Observation 3: In our simulation, RS overhead reduction is up to 87.5% with at least around 76% and 79% Top-1 accuracy of beam prediction.

NVDIA [12]:

* Observation 2: AI/ML-based algorithms for beam prediction in spatial domain can achieve performance comparable to that of exhaustive beam search, while the reference signal overhead, measurement effort, reporting overhead, and latency can be much reduced.
* Proposal 6: Capture the presented evaluation results in the TR to highlight that AI/ML-based algorithms for beam prediction in spatial domain can achieve performance comparable to that of exhaustive beam search while reducing the reference signal overhead, measurement effort, reporting overhead, and latency.

Xiaomi [14]:

* Observation 1: AI based beam prediction in spatial domain can provide good performance. And the performance can be further improved by inputting corresponding beam pair ID in addition to measured L1-RSRP or by inputting L1-RSRP of same beam pair IDs.

China Telecom [16]:

* Observation 1: Modelling the spatial beam prediction task as a classification model provides better performance with less training overhead.

Samsung [17]:

* Observation # 3: For spatial domain prediction, AI can provide better performance in terms of beam prediction accuracy than non-AI based scheme with the measurements of a given subset of beams to select a best beam among a full set of beams.
* Observation # 4: With the help of AI, SSB/RS overhead for measurements, UE measurement efforts, reporting overheads can be reduced to achieve a target performance for beam selection.
* Observation # 5: For spatial domain prediction, AI can provide better performance in terms of beam prediction accuracy than non-AI based scheme with the measurements of a set of wide beams and a subset of narrow beams to select a best beam among a full set of narrow beams.
* Observation # 6: For spatial domain prediction, AI can predict the best narrow beam based on the measurements of wide beams only with decent performance.
* Observation # 7: For spatial domain prediction, AI can help gNB to predict the best narrow beam set that including the best narrow beam for UE to measure with high probability.

CMCC [19]

* Observation 1: The increase of K significantly improves the prediction accuracy while leading to a small degree of increased beam sweeping overhead.
* Observation 2: Compared with baseline option 1, AI based spatial beam prediction has minor loss of prediction accuracy for top-K beam pair but has large beam sweeping overhead reduction.
* Observation 3: Compared with baseline option 2, AI based spatial beam prediction significantly enhances prediction accuracy for top-K beam pair under the same beam sweeping overhead.

Ericsson [20]

* Observation 5 In outdoor scenarios, AI/ML can reduce beam spatial-domain beam prediction overhead substantially while maintaining good accuracy, both for 4x8 (30 beams in Set A) and 8x16 arrays (168 beams in Set A).
* Observation 6 In scenarios with primarily indoor UEs, spatial-domain beam predication is more challenging.
* Observation 7 In scenarios with primarily indoor UEs, spatial-domain beam predication is more challenging, but increased NN complexity is found to improve performance.

Nokia [21]

* Observation 3: For BM-Case1, Set B RSRP may not be sufficient for beam prediction input in certain cases.
* Observation 4: For BM-Case1, the beam prediction failure detection procedure is needed to be further studied.
* Observation 13: For Set B is different to Set A with Set B is wide beam, the KPI for the wide beam codebook design should be both prediction accuracy and throughput performance.
* Observation 15: Selecting the beam based on the RL agent output can improve the throughput performance of each UE by clustering the UEs to a single beam. However, the gains at different percentiles depend on the data rate and channel bandwidth.

MediaTek [22]:

* Observation 3: With a greater number of beams in Set B, both models achieve higher top-k accuracy. However, greater number of beams in Set B requires more beam RSRP measurements.
* Observation 6: The spatial beam prediction by using wide beam design in Set B does not outperforms the performance by using subset beam design in Set B.

CEWiT [23]

* Observation 1: When compared to the baseline scheme, the RS overhead of the AI/ML model can be reduced by 50% with 0.37dB RSRP difference and a Top-1 accuracy of 74.47%.
* Observation 2: When compared to the baseline scheme, the RS overhead of the AI/ML model can be reduced by 75% with 0.78dB RSRP difference and a Top-1 accuracy of 70.63%.
* Observation 3: When compared to the baseline scheme, the RS overhead of the AI/ML model can be reduced by 87.5% with 1.68dB RSRP difference and a Top-1 accuracy of 65.02%.

Apple [25]

* Observation 1: For case 1, AI based beam search space prediction can provide up to 30% gain compared to non-AI based approach.

DoCoMo [26]

* Observation 2: Spatial domain DL beam prediction could notably improve the beam prediction accuracy.

## 3.2 Temporal domain prediction (on hold)

The following observations were provided in contributions:

Huawei [2]:

* Observation 5: For temporal beam prediction, AI/ML based methods are more robust than legacy approaches to variations of the UE speed.
  + When the time duration is 0.08s in the observation and prediction window, for UE speed 30km/h, the AI/ML Top-8 approach is 42% better than for the legacy baseline but for a UE speed of 90 km/h, the AI/ML Top-8 prediction accuracy is 47% better than for the legacy baseline
  + When the time interval is 0.16s in the observation and prediction window, for UE speed 30km/h, the AI/ML Top-8 approach is 48% better than for the legacy baseline but for UE speed 90 km/h, the AI/ML Top-8 prediction accuracy is 77% better than for the legacy baseline.
* Observation 6: For temporal beam prediction, lower spatial consistency has more impact on the prediction accuracy achieved by the legacy approach than on accuracy achieved by the AI/ML-based methods. This can be seen from the results when different time durations are evaluated.
  + For UE at 30km/h, the accuracy of AI/ML Top-8 degrades 3.35% but the baseline degrades 4.8% when stretching the two prediction durations from 0.08s to 0.16s
  + For UE at 90km/h, the accuracy of AI/Ml Top-8 degrades 0.93% but the baseline degrades 9.56% when stretching the two prediction durations from 0.08s to 0.16s

Vivo[3]:

* For BM-case2, compared with non-AI scheme, beam pair prediction scheme improves beam prediction accuracy and reduces average L1-RSRP difference significantly.

ZTE [4]

* Observation 6: Compared with the selected non-AI method, a better performance can be obtained by the AI based temporal beam prediction method in terms of beam prediction accuracy for Top-1 and Top-4 beams.
* Proposal 7: The AI/ML model can be utilized for temporal beam prediction especially for a large sampling interval or UE speed.

OPPO [9]

* Observation 5: Temporal domain beam prediction can provide good beam prediction accuracy (at least 77%) while overhead/latency reduction can be up to 50% (for the case of K = 4 and F = 4).
* Observation 6: Beam predication accuracy slightly decreases from 87.1% to 77.1% (the case of Top-1) when F increases from 1 to 4, but strongly increases from 77.1% to 98.8% (the case of F = 4) when predicted beam number increases from Top-1 to Top-4.
* Observation 7: For 80% of the incorrect temporal domain beam prediction cases, the L1-RSRP difference is lower than 3.5dB which may not strongly impact the spectrum efficiency.

NVDIA [12]

* Observation 3: AI/ML-based algorithms for beam prediction in time domain can simply use a history of the best beam index to perform the prediction.
* Observation 4: AI/ML-based algorithms for beam prediction in time domain can help lower reference signal overhead and reduce UE’s measurement requirement.
* Proposal 7: Capture the presented evaluation results in the TR to highlight that AI/ML-based algorithms for beam prediction in time domain can simply use a history of the best beam index to perform the prediction to lower reference signal overhead and reduce UE’s measurement requirement.

Xiaomi [14]:

* Observation 2: AI based beam prediction scheme 1 and scheme 2 in time domain can provide good performance.
  + Scheme 1 assumes same periodicity for history measurement instance and future time instance.
  + Scheme 2 assumes longer periodicity for history measurement instance than that of future time instance. It can reduce more RS overhead than scheme 1.
* Observation 3: Set B < set A causes much more performance degradation compared to set B=set A for temporal beam prediction.
* Observation 4: The performance may degrade when larger N (history measurement instances) is assumed.

Samsung [17]:

* Observation # 8: For time and spatial domain prediction, AI can provide better performance in terms of beam prediction accuracy than non-AI based scheme with the measurements of a subset of narrow beams to select a best beam among a full set of narrow beams.

MediaTek [22]:

* Observation 1: Both machine learning models perform better on ray-tracing dataset compared to SLS dataset.
* Proposal 4: Evaluate the impact of different observation and prediction window sizes to the performance of AI/ML temporal beam prediction.

DoCOMo [26]:

* Observation 3: For the UE speed 30km/h, baseline performance is nice enough even if 960ms reporting periodicity is assumed.
* Observation 4: For the UE speed 60km/h and higher, the AI/ML model could provide notable improvement in prediction accuracy.

# Others

Some companies suggest to consider multiple scenarios for evaluations.

* NVIDIA [12]: Proposal 1: Beam prediction in spatial domain and beam prediction in time domain should be the focal point for evaluating AI/ML based algorithms for beam management.
  + FL1: Can be discussed in 9.2.3.2
* Intel [13]: Proposal 1: For AI/ML evaluation for beam management use cases, including spatial and temporal domain beam management, consider only offline training of AI/ML models.
  + FL1: Can be discussed in 9.2.3.2
* Qualcomm [24]
  + Proposal 5: Study the benefits and trade-offs associated with UE-side and gNB-side temporal beam prediction, using the agreed KPIs
  + Proposal 9: RAN1 should study the benefits and trade-offs associated with UE-side and gNB-side spatial (+time) domain beam prediction, using the agreed KPIs
  + FL1: Can be discussed in later stage

# Reference

[1] [R1-2205753](file:///C:\Users\younsun\Documents\3GPP%20documents\RAN1%20tdocs\TSGR1_110\Docs\R1-2205753.zip) Continued discussion on evaluation of AI/ML for beam management FUTUREWEI

[2] [R1-2205892](file:///C:\Users\younsun\Documents\3GPP%20documents\RAN1%20tdocs\TSGR1_110\Docs\R1-2205892.zip) Evaluation on AI/ML for beam management Huawei, HiSilicon

[3] [R1-2206034](file:///C:\Users\younsun\Documents\3GPP%20documents\RAN1%20tdocs\TSGR1_110\Docs\R1-2206034.zip) Evaluation on AI/ML for beam management vivo

[4] [R1-2206070](file:///C:\Users\younsun\Documents\3GPP%20documents\RAN1%20tdocs\TSGR1_110\Docs\R1-2206070.zip) Evaluation on AI for beam management ZTE

[5] [R1-2206166](file:///C:\Users\younsun\Documents\3GPP%20documents\RAN1%20tdocs\TSGR1_110\Docs\R1-2206166.zip) Evaluation on AI/ML for beam management Fujitsu

[6] [R1-2206181](file:///C:\Users\younsun\Documents\3GPP%20documents\RAN1%20tdocs\TSGR1_110\Docs\R1-2206181.zip) Discussion for evaluation on AI/ML for beam management InterDigital, Inc.

[7] [R1-2206197](file:///C:\Users\younsun\Documents\3GPP%20documents\RAN1%20tdocs\TSGR1_110\Docs\R1-2206197.zip) On Evaluation of AI/ML based Beam Management Google

[8] [R1-2206250](file:///C:\Users\younsun\Documents\3GPP%20documents\RAN1%20tdocs\TSGR1_110\Docs\R1-2206250.zip) Evaluation of AI/ML based beam management Rakuten Mobile, Inc

[9] [R1-2206317](file:///C:\Users\younsun\Documents\3GPP%20documents\RAN1%20tdocs\TSGR1_110\Docs\R1-2206317.zip) Evaluation methodology and preliminary results on AI/ML for beam management OPPO

[10] [R1-2206393](file:///C:\Users\younsun\Documents\3GPP%20documents\RAN1%20tdocs\TSGR1_110\Docs\R1-2206393.zip) Evaluation on AI/ML for beam management CATT

[11] [R1-2206512](file:///C:\Users\younsun\Documents\3GPP%20documents\RAN1%20tdocs\TSGR1_110\Docs\R1-2206512.zip) Evaluation on AI/ML for beam management Lenovo

[12] [R1-2206522](file:///C:\Users\younsun\Documents\3GPP%20documents\RAN1%20tdocs\TSGR1_110\Docs\R1-2206522.zip) Evaluation of AI and ML for beam management NVIDIA

[13] [R1-2206580](file:///C:\Users\younsun\Documents\3GPP%20documents\RAN1%20tdocs\TSGR1_110\Docs\R1-2206580.zip) Evaluation for beam management Intel Corporation

[14] [R1-2206637](file:///C:\Users\younsun\Documents\3GPP%20documents\RAN1%20tdocs\TSGR1_110\Docs\R1-2206637.zip) Evaluation on AI/ML for beam management Xiaomi

[15] [R1-2206677](file:///C:\Users\younsun\Documents\3GPP%20documents\RAN1%20tdocs\TSGR1_110\Docs\R1-2206677.zip) Some discussions on evaluation on AI-ML for Beam management CAICT

[16] [R1-2206688](file:///C:\Users\younsun\Documents\3GPP%20documents\RAN1%20tdocs\TSGR1_110\Docs\R1-2206688.zip) Evaluation on AI/ML for beam management China Telecom

[17] [R1-2206822](file:///C:\Users\younsun\Documents\3GPP%20documents\RAN1%20tdocs\TSGR1_110\Docs\R1-2206822.zip) Evaluation on AI ML for Beam management Samsung

[18] [R1-2206876](file:///C:\Users\younsun\Documents\3GPP%20documents\RAN1%20tdocs\TSGR1_110\Docs\R1-2206876.zip) Evaluation on AI/ML for beam management LG Electronics

[19] [R1-2206904](file:///C:\Users\younsun\Documents\3GPP%20documents\RAN1%20tdocs\TSGR1_110\Docs\R1-2206904.zip) Discussion on evaluation on AI/ML for beam management CMCC

[20] [R1-2206938](file:///C:\Users\younsun\Documents\3GPP%20documents\RAN1%20tdocs\TSGR1_110\Docs\R1-2206938.zip) Evaluation on AI/ML for beam management Ericsson

[21] [R1-2206970](file:///C:\Users\younsun\Documents\3GPP%20documents\RAN1%20tdocs\TSGR1_110\Docs\R1-2206970.zip) Evaluation of ML for beam management Nokia, Nokia Shanghai Bell

[22] [R1-2206990](file:///C:\Users\younsun\Documents\3GPP%20documents\RAN1%20tdocs\TSGR1_110\Docs\R1-2206990.zip) Evaluation on AI/ML for beam management MediaTek Inc.

[23] [R1-2207068](file:///C:\Users\younsun\Documents\3GPP%20documents\RAN1%20tdocs\TSGR1_110\Docs\R1-2207068.zip) Evaluation on AI/ML for beam management CEWiT

[24] [R1-2207226](file:///C:\Users\younsun\Documents\3GPP%20documents\RAN1%20tdocs\TSGR1_110\Docs\R1-2207226.zip) Evaluation on AI/ML for beam management Qualcomm Incorporated

[25] [R1-2207330](file:///C:\Users\younsun\Documents\3GPP%20documents\RAN1%20tdocs\TSGR1_110\Docs\R1-2207330.zip) Evaluation on AI/ML for beam management Apple

[26] [R1-2207403](file:///C:\Users\younsun\Documents\3GPP%20documents\RAN1%20tdocs\TSGR1_110\Docs\R1-2207403.zip) Discussion on evaluation on AI/ML for beam management NTT DOCOMO, INC.

# Appendix: Agreements

# Agreements in RAN 1 #109e

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

From May 17th GTW session

Agreement

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

Agreement

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

Agreement

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

Agreement

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

Agreement

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

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

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

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

Agreement

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

Agreement

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

Conclusion

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

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

Agreement

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

Agreement

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

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

Agreement

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

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

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

Agreement

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



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

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

For training data generation

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

Agreement

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

Agreement

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

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

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

Agreement

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

Agreement

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

Agreement

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

Final summary in R1-2205641.