**3GPP TSG RAN WG1 #109-e R1-220xxxx**

**e-Meeting, May 09 – May 20, 2022**

**Agenda item:** **9.2.3.1**

**Source:** Moderator (Samsung)

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

**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 0-1a

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

|  |  |  |
| --- | --- | --- |
| Company | Point of contact | Email address |
| Apple | Yushu Zhang | yushu\_zhang@apple.com |
| Nokia, NSB | Keeth Jayasinghe, Mihai Enescu | [keeth.jayasinghe@nokia.com](mailto:keeth.jayasinghe@nokia.com), mihai.enescu@nokia.com |
| Xiaomi | Mingju Li | limingju@xiaomi.com |
| vivo | Peng SUN | sunpeng@vivo.com |
| Intel | Avik Sengupta | avik.sengupta@intel.com |
| NVIDIA | Xingqin Lin | xingqinl@nvidia.com |
| OPPO | Zhihua Shi | szh@oppo.com |
| AT&T | Thomas Novlan | thomas\_novlan@labs.att.com |
| CATT | Yongqiang FEI | feiyongqiang@catt.cn |
| LG Electronics | SeongWon Go | [sw.go@lge.com](mailto:sw.go@lge.com) |
| Ericsson | Henrik Ryden | Henrik.a.ryden@ericsson.com |
| ZTE, Sanechips | Wenfeng Liu | liu.wenfeng@zte.com.cn |
| CAICT | Xiaofeng Liu | Liuxiaofeng1@caict.ac.cn |
| Samsung | Jackson Wang  Yeon-Geun Lim | [h0809.wang@samsung.com](mailto:h0809.wang@samsung.com)  yg.lim@samsung.com |
| Fujitsu | Wang Xin | wangxin@fujitsu.com |
| CMCC | Jun ZUO | zuojun@chinamobile.com |
| MediaTek | Gyu Bum Kyung | gyubum.kyung@mediatek.com |
| Huawei, HiSilicon | Thorsten Schier | Thorsten.schier@huawei.com |
| IDCC | Youngwoo Kwak | youngwoo.kwak@interdigital.com |
| Lenovo | Srinivas Kothapalli | vkothapalli@lenovo.com |
| Qualcomm | Hamed Pezeshki | hamedp@qti.qualcomm.com |

# Evaluation methodology on AI/ML in beam management

## 1.1 Dataset construction

Most of the companies proposed to use the dataset generated based on statistical channel models for evaluation:

* Huawei/HiSi [1]: For AI/ML-based beam management evaluation, training inputs generated from simulation platform should be a baseline.
* ZTE [2]: Since the data sets and AI models used by different companies are different, it is necessary to provide common data sets and baseline models for simulation calibration and performance cross-validation.
* Ericsson [4]: Beam management should be evaluated based on the scenarios defined in (simulation assumption of Rel-17 BM in R1-2007151)
* CATT [6]: A few typical scenarios in TR 38.901 can be considered for the evaluations for AI/ML-based beam management.
* vivo [7]: Data set constructed based on 3GPP channel model from TR 38.901 is mainly considered for beam management case. SLS deployment is used to obtain dataset for performance evaluation.
* Apple [15]: System level simulator can be used to generate the dataset, where the number of cells can be 1 to speed up the dataset generation procedure and full spatial consistency modeling is recommended.
  + Other parameters can be selected based on traditional system level evaluation
* DCM [17]: 3GPP statistical channel models are considered in the evaluation for representative sub use-case selection.
* DCM [17]: Discuss and decide whether and which deterministic channel models should be used to capture the final evaluation results of selected sub use-cases.
  + Option 1: Field data as indicated in SID.
  + Option 2: Ray-tracing channel model.
  + Option 3: Map-based hybrid channel model as defined in TR38.901.
* Lenovo [18]: Generate the dataset for the model training and evaluation for beam management sub use cases based on the statistical models in TR 38.901 taking in to account spatially consistency.
* Nokia/NSB [19]: The preliminary phase uses synthetic data generated from a system-level simulator that implements the 3GPP statistical channel models from 3GPP TR 38.901. The list of assumptions detailed in Table 2.1 2 can be considered for system-level simulations of different beam management sub-use cases.
* NVIDIA [21]: Use the simulation assumptions in TR 38.901 as a starting point for the evaluation of AI/ML based algorithms for beam management.
* Qualcomm [23]: RAN1 to use channel model in TR 38.901 for the evaluations on AI/ML for temporal beam prediction. A set of simulation assumptions should be agreed upon for evaluation purposes.
* Qualcomm [23]: RAN1 to use channel model in TR 38.901 for the evaluations on AI/ML for codebook-based spatial domain beam prediction. A set of simulation assumptions should be agreed upon for evaluation purposes.

Some companies suggested to encourage companies to provide common/public dataset for training and validation for cross check purposes. In [14], a website for mobile communication open dataset was mentioned as <http://www.mobileai-dataset.com/>.

* Vivo [7]: It is encouraged for companies to provide publicly accessible dataset for training and validation for cross-check purposes.
* Intel [20]: A common dataset across companies should be considered for each use-case to ensure robustness and fair comparison of AI/ML model performance taking into account, a reasonable dataset size.

Moreover, one company proposed that no need to use field data for the study item beam management use case, while two companies proposed that real data should be part of the evaluation work, and suggested to encourage companies to contribute realistic data to develop and evaluation.

* Ericsson [4]: No need for field data for the study item BM use case
* DCM [17]: Discuss and decide whether and which deterministic channel models should be used to capture the final evaluation results of selected sub use-cases.
  + Option 1: Field data as indicated in SID.
  + Option 2: Ray-tracing channel model.
  + Option 3: Map-based hybrid channel model as defined in TR38.901
* NVIDIA [21]: Identifying existing sets of real data should be part of the evaluation work for AI/ML based beam management.
* NVIDIA [21]: Companies are encouraged to contribute real data to develop and evaluate AI/ML based algorithms for beam management.

Based on the above views, the following proposal related to dataset construction can be considered.

#### FL1 High Priority Question 1-1

**Proposal 1-1:**

* **Dataset constructed based on 3GPP channel model(s) in TR 38.901 is used for beam management use case.**

**Question 1-1:**

1. Whether the above proposal 1-1 can be adopted?
2. Whether real data/field data is optionally needed as part of the study?

|  |  |  |
| --- | --- | --- |
| Company | Y/N | Comments |
| Apple | Y |  |
| Nokia, NSB | Y | Yes, Support a)  Optional field data should be fine but should not be the main focus. |
| Xiaomi | Y | Support a) |
| vivo | Y | At least in SI stage, no need to introduce real data/field data |
| Intel | Y | Statistical channel models are OK as long dataset generation assumptions can be well aligned across companies |
| NVIDIA | Y | Support a)  Support b) to consider optional field data. |
| OPPO | Y | The real data / field data are welcome and can be used optionally. |
| AT&T | Y |  |
| CATT | Y | Support a).  Field data can be optionally used, if companies have interest. |
| LGE | Y | a) Support.  b) We think stochastic channel model is enough to evaluate AI/ML for BM. |
| Ericsson | Y | Support a) |
| ZTE, Sanechips | Y | a) Support.  b) Field data is optionally needed but should be placed at a lower priority. |
| CAICT | Y | Field data is not necessary for sub case down selection at least. |
| Samsung | Y | Yes for a)  For b), no need to spend time on field data at this stage. |
| Fujitsu | Y | It’s not necessary to use the real data/field data in initial stage. |
| CMCC | Y |  |
| MediaTek | Y | In SI phase, we can use dataset generated by 3GPP model(s) |
| HW/HiSi | Y | a) Support  b) Field data is optional needed and should not be mandatory. |
| InterDigital | Y | a) We support using 3GPP channel model(s) for the evaluation.  b) We are fine with having real/field data if it is optional. |
| Lenovo | Y | 1. Yes 2. Open to the idea of having field data as optional. Further discussions are needed to decide on collecting the field data. |
| Qualcomm | Y | 1. We believe TR 38.901 encapsulates the necessary tools (e.g., spatial consistency) for the use cases under consideration. 2. No need for optional real data/field data. |

## 1.2 Dataset generation and evaluation assumptions with SLS

Most of companies proposed to use SLS to evaluate the performance of beam management. Two companies explicitly proposed to use evaluation methodology (EVM) for Rel-17 beam management (R1-2007151), while other companies provided detailed proposals and analysis for the evaluation assumption for SLS. Table 1-1 in appendix summarized the main parameters proposed/used in some contributions.

* Ericsson [4]: Beam management should be evaluated based on the scenarios defined in (simulation assumption of Rel-17 BM), which are summarized in the appendix
* Samsung [9]: Consider Table A.2.5-2 (for SLS) and Table A.1.6.4 (for LLS) of TR 38.802 as the starting point for Rel-18 AI/ML for beam management evaluation, based on which necessary modifications/simplifications can be further discussed for the specific evaluation purpose and/or sub use case.
* LGE [13]: Reuse agreed simulation assumptions in Rel-17 BM (R1-2007151) except AI/ML related function for Rel-18 evaluation on AI/ML for beam management.
* CAICT [14]: Umi scenario with carrier frequency 30GHz could be considered as baseline for dataset construction and performance comparison.
* Nokia/NSB [19]: The preliminary phase uses synthetic data generated from a system-level simulator that implements the 3GPP statistical channel models from 3GPP TR 38.901. The list of assumptions detailed in Table 2.1 2 can be considered for system-level simulations of different beam management sub-use cases.
* Intel [20]: For temporal domain beam prediction, system level simulation based urban deployment e.g., dense urban and urban macro and related 3GPP channel models should be used for dataset generation
* NVIDIA [21]: The evaluation scenarios for beam management include UMi-street canyon and UMa scenarios.
* NVIDIA [21]: Focus on millimeter wave frequencies for the evaluation of AI/ML based algorithms for beam management.
* AT&T [22]: Consider both a single-cell and multi-cell SLS evaluation scenarios
* AT&T [22]: Consider modified Urban Macro and Indoor Hotspot layouts with non-uniform inter-site distances and antenna height/orientations
* AT&T [22]: Consider a venue/stadium layout with a dense deployment of base stations/TRPs oriented towards one or more cluster of users

Based on the above views, the following proposal related to evaluation methodology can be considered. Moreover, Table 1 in Appendix summarized and simplified the baseline assumptions Table 1 and Table 2 in R1-2007151 for Rel-17 BM EVM. The parameters for high-speed train scenario were removed comparing with the tables in R1-2007151, which will be separated discussed in Question 1-9. Companies are encouraged to indicate necessary changes of those parameters and the reasons.

#### FL1 High Priority Question 1-2

**Proposal 1-2:**

* **For dataset construction and performance evaluation for AI/ML based BM, SLS is the baseline tool.**

**Question 1-2:**

1. Whether the above proposal 1-2 can be adopted?
2. Which parameters in Table 1 on baseline assumptions for SLS need to be modified, and why?
3. Whether Dense Urban scenario (FR2) can be used as the baseline assumption for AI/ML for BM? If the answer is yes, which parameters in Table 2 need to be updated, and why?
4. Any other scenario(s) can be optionally considered for dataset generation and performance evaluation of AI/ML for BM?

Note: Mixed scenario(s) for generalization can be discussed in section 2.2.1. This question only focuses on single scenario case for dataset and performance evaluation.

|  |  |  |
| --- | --- | --- |
| Company | Y/N | Comments |
| Apple | Y | For simulation assumption, we think we should add FR1 related parameters since one use case is to use FR1 to predict FR2 beam. For scenario, we think 1 cell with 3 sectors can be added to be aligned with intra-cell mobility assumption in R17. For UE speed, we suggest adding 30 km/h. |
| Nokia, NSB | Y | b.) “One UE is dropped for each of the 21 sectors/cells"., we do not support this assumption for the reasons as follows:   * it is an oversimplified scenario that would complicate the data generation as the data collected for each network drop is limited, * SLS perf evaluation is not realistic as the single UE assumption does impact the interference and consequently the throughput, which is likely to provide unreal results. * At least 10 UEs/sector dropped uniformly in random points of the sector should be considered.   Also, “UE Speed 120 km/h” is not realistic for Dense Urban scenario. Consider including also 30/45/60/75/90 km/h. These speeds could also be sub-use case specific.  c.) Yes.  The scenario "UMa LOS" is too simplistic as building and other static street objects may obstruct the BS-UE link. We suggest that for a given link, the models decide whether the channel propagation conditions are LoS or NLoS, by considering a distance-dependent LoS probability function as currently done in 38.901.  d.) yes, we suggest considering other options for list of assumptions in [19] (detailed in Table 2.1 2 (right hand side column)) for beam prediction in spatial domain for throughput & latency reduction. |
| Xiaomi | Y | First we want to clarify that the parameters in Table 1 will be used for beam prediction in both spatial domain and time domain?  Then for b), we also think that based on “One UE is dropped for each of the 21 sectors/cells”, it is not better for data collection. For UE speed, we prefer to add 30km/h. |
| vivo | Y | We prefer 80% indoor and 20% outdoor for spatial domain beam prediction and 100% outdoor for temporal domain prediction.  For UE speed, 120km can be optional. 30km/h should also be included.  For gNB and UE antenna configuration, single panel can also be optional configurations to reduce the simulation time and effort in addition to multiple panel simulations.  The number of Tx beams and Rx beams should be aligned, and we prefer at least a total of 256 beam pairs should be used. |
| Intel | N | Q1-2a: In our understanding, LLS based channel models are good enough for spatial domain beam prediction and SLS based data is needed only for temporal domain beam prediction use case. Therefore, the baseline assumption should be sub-use-case specific.  Q.1-2b: Considering Table 1, the UE dropping should be revisited along with trajectory modeling and UE speed. Rel-17 MIMO EVM was for mobility with highway traffic. In temporal beam prediction, the use case should be further refined, and highway traffic may not be the only scenario of interest. A more general assumption of random UE dropping and straight-line motion at a lower speed can be considered to collect traces of UE beam switching data for temporal prediction.  For both the remaining questions, the assumptions should be sub-use-case specific and Dense Urban for example should be relevant only for the temporal domain prediction. |
| NVIDIA | Y | Support Proposal 1-2 |
| OPPO | Y | a) Yes  b) For the sub use case of beam prediction in spatial domain, it seems not mandatory for UE to move at high speed e.g. 60km/h or even higher speed. When UE travels at low speed, e.g. 3km/h, the beam prediction in spatial domain can be evaluated.  c) Indoor hotspot (office scenario) also serves as classic FR2 deployment and should be considered for dataset generation and evaluation for spatial domain prediction. |
| AT&T | Y | d) A stadium/venue scenario could optionally be considered as well as the indoor hotspot deployment scenario - both for FR2 |
| CATT | Y | It is good to focus on FR2 only. Additionally, AI-based beam management seems more attractive in outdoor due to the challenge brought by high speed. Indoor scenario may not need AI-based approach urgently.  For b), we are open to further consider additional slower UE speed(s), e.g. 3km/h, 30km/h.  For traffic model, FTP traffic shall be evaluated. |
| LGE | Y | a) Support.  b) Agree with adding 3 km/h, 30 km/h for UE speed.  c, d) OK. sub use case-specific scenarios can be optionally considered and companies can provide detailed assumption. |
| Ericsson | Y | 1. Agree 2. -We encourage the possibility of using other traffic models than full buffer. Full buffer cannot be used to make strong predictions of the usefulness of a feature and hence doesn’t provide any insights that assist in drawing SI conclusions.   -Other UE speeds than 60/120 km/h should also be simulated given the Dense Urban scenario. Such as 3/10/30 km/h |
| ZTE, Sanechips | Y | We generally think that the baseline SLS assumptions should be sub-use-case specific to distinguish the spatial-domain and time-domain beam prediction. The following revisions are suggested to be considered further.   1. UMi scenario for spatial-domain beam prediction and HST scenario for time-domain beam prediction can be optionally simulated. 2. Multiple UEs can be dropped for each of the 21 sectors/cells for AI/ML model generalization improvement. 3. As mentioned by some other companies, more UE mobility such as 3km/h, 30km/h and 300km/h can be included for evaluation. 4. An UE antenna configuration of 2 panels (left, right) can also be taken into consideration especially for spatial-domain beam prediction. As captured in Table A.2.1-4 of TR 38.802, a suitable configuration option of UE antenna is provided as (Mg, Ng) = (1, 2) for NR MIMO evaluation. |
| CAICT | Y | SLS for FR2 with single scenario case for dataset should be baseline. |
| Samsung | Y | 1. Yes. Although we think LLS can be used for time domain beam prediction, at this stage we think LLS should also be considered, especially considering some of LLS is effective for some use cases of time domain prediction. 2. For Table 1, the following parameters can be further considered:   - Same as comments from Apple and others, 30kmph can be a good starting point to be researched.  - For panel number, as indicated in our paper, the simplified BS antenna configuration with only 1 panel used can be considered, which is equivalent in beam selection accuracy performance to be evaluated in Phase-I.  - For beam correspondence, to simplify the study, we can prioritize BC without SRS assistance, which is more common in existing product.   1. Yes. Parameters in Table 2 is good to us.   Other scenarios can be considered if the existing SLS setup is not appropriate for a certain use case which is accepted in the use case email thread. |
| Fujitsu | Yes | 1. Yes. 2. The beam prediction in spatial domain and time domain should be provided with different parameter tables.   In table 1, 120km/h in urban scenario is not practical, UE mobility with 30km/h and 90km/h are recommended.  For spatial domain prediction, at least the following parameters are recommended   * Low UE mobility (e.g. 3km/h)   more UE drops per sector/cell (e.g. 10 dropped UEs per sector/cell) for easy training data collection. |
| CMCC | Y | a) Besides SLS, LLS based channel model can be optional used.  b) and d) We propose to consider two scenarios. One is dense urban with low UE speed (e.g., 30km/h, 60km/h), another is HST/highway scenario with UE speed 120km/h, 350km/h, 500km/h. |
| MediaTek | Y | b. The number of UEs dropped in each cell needs not to be limited. Also, lower UE speeds may need to be considered (e.g., 3km/h and 30km/h)  c. We agree that FR2 can be used as the baseline assumption |
| HW/Hisi | Y | a) LLS should also be considered at least for spatial domain prediction.  b) For table 1, the following modification can be considered:   * BW can be 20MHz for simplifying simulation and accelerating model training at the beginning stage * 3km/h , 30km/h can be added to UE speed * For gNB and UE antenna configuration, single panel can be optional to reduce the simulation time and this would also show the performance gain.   c) yes, NLOS should also be considered. |
| InterDigital | Y | a) We agree that system level simulation should be a baseline.  b) 1. We believe that the suggest evaluation assumptions are focused on L1-RSRP evaluation with time domain prediction, but includes some other evaluation assumptions for system level performance and spatial domain prediction. We suggest to provide different evaluation assumptions for each scenario reflecting our comments in the below.  2. One UE dropped for each cell may be fine for L1-RSRP measurement but not fine for evaluating system performance. We suggest to include 10 UEs for evaluation system performance.  3. 120 km/hr in dense urban is not a realistic evaluation assumption. We suggest to add 3km/hr and 30 km/hr.  4. Link adaptation, traffic model and UE receiver type are for evaluating system level performance, but system level performance is not included in ‘Algorithm details (when applicable)’. We suggest adding system level performance.  5. For evaluating system level performance, full buffer traffic model is not realistic. We suggest using FTP model instead of full buffer traffic.  c) We are fine with Dense Urban scenario (FR2).  d) For spatial domain prediction, view, indoor channels such as indoor factory should be included as well as dense urban. For indoor scenario, we suggest the following antenna configuration.  Panel model 1, Mg = 1, Ng = 1, P = 2, dH = 0.5λ, (M, N, P, Mg, Ng) = (2, 4, 2, 1, 1) |
| Lenovo | Y | 1. Yes 2. Along with `High Speed@FR2’, it is recommended to also consider moderate and low speeds, for example, including 3km/hr and 30km/hr for UE Speed for spatial domain beam prediction. 3. Yes. 4. Consider defining a set of multiple scenarios/configurations, such as low/moderate UE speeds, Different antenna configurations (with more than one antenna panel) at gNB/UE. It is also recommended to define baseline EVM that can evaluate the generalization and scalability of the proposed models. |
| Qualcomm |  | 1. Yes 2. 1- “One UE is dropped for each of the 21 sectors/cells” 🡪 In this case, we need to make sure the training dataset is generic enough. To achieve this, we can consider sufficient number of random initial UE deployment in each cell and random UE trajectory, as discussed in Option #2 and Option #3 in Section 1.2.2 of this document. Considering random UE trajectory, the prospect of moving UEs across different cells need to be considered, as it will affect the beam prediction performance.   2- More set of UE speeds should be added including but not limited to 30 km/h. To evaluate the generalization capability of AI/ML model, it would be beneficial to have the option to train and test over different UE speeds.   1. Yes, but do not see the necessity to consider “UMa LOS” as the baseline scenario. Baseline scenario could be UMa with mixed LOS/NLOS. 2. Indoor hotspot can also be considered as a scenario with 3km/h UE speeds. The prospects of random UE trajectory and therefore moving across cells need to be considered. |

### 1.2.1 Spatial consistency

Several companies discussed spatial consistency for UE time domain beam prediction.

* Vivo [7]: Spatial consistency should be modeled for UE mobility scenario for time domain beam prediction.
* Samsung [9]: For evaluation of beam prediction in the time domain, in addition to the spatially-consistent mobility modelling and UE trajectory modelling, the selection of UE speed, blockage [Section 7.6.4 in TR 38.901], and/or UT rotating [Section 7.6.7 in TR 38.901] should be considered.
* Samsung [9]: Spatially-consistent UE mobility modelling: Procedure A in clause 7.6.3.2 in TR38.901
  + Note: Spatial consistency model is needed to evaluate beam prediction performance, by generating the correlated cluster power, delay, AOD/AOA/ZOD/ZOA.
* Futurewei [12]: To properly model and assess the performance of prediction in time domain, spatial consistency along the mobility trajectory needs to be ensured.
* Apple [15]: System level simulator can be used to generate the dataset, where the number of cells can be 1 to speed up the dataset generation procedure and full spatial consistency modeling is recommended.
  + Other parameters can be selected based on traditional system level evaluation
* Lenovo [18]: Generate the dataset for the model training and evaluation for beam management sub use cases based on the statistical models in TR 38.901 taking in to account spatially consistency.
* Nokia/NSB [19]: A suitable UE mobility model is needed for the spatial-temporal domain beam prediction study.
  + As shown in table 2.4-1, the current UE mobility model for spatial-temporal beam prediction assumes a straight trajectory traveled by the UE with a random direction and channel spatial consistency type-A defined in TR 38.901.
* Intel [20]: Spatially consistent large-scale parameter generation should be used for mobility evaluations. Additionally, only spatial consistency model B in [TR 38.901] can be used for mobility evaluation.

Based on the above views, the following proposal related to spatial consistency for mobility evaluation can discussed.

#### FL1 High Priority Question 1-3

**Question 1-3:**

1. Whether spatial consistency should be modeled for time domain beam prediction?
2. Which procedure can be used, procedure A or procedure B in TR38.901, and why?

|  |  |
| --- | --- |
| Company | Comments |
| Apple | a) Y. We think it is also needed for spatial domain beam prediction  b) We think cluster level spatial consistency defined in 7.6.3.1 in 38.901 is enough. |
| Nokia, NSB | 1. Yes, since spatial consistency enables to capture the correlated behaviour of the channel for UE moving over a path 2. We support the use of spatial consistency procedures defined in TR38.901. |
| vivo | a) Support spatial consistency for time domain prediction  b) section 7.6.3.1 in 38.901 is sufficient |
| Intel | Spatial consistency should be modeled and as explained in our paper, Model B in TR38.901 is relevant for mobility where the UE moves across the deployment in a trajectory. |
| NVIDIA | 1. Yes, it’s necessary to model spatial consistency 2. Though the use of spatial consistency procedures defined in TR38.901 can be the baseline, it would be good to check its modeling accuracy. |
| OPPO | a) Yes  b) We are open to procedure A or B |
| AT&T | 1. Yes, needed for spatial and time domain 2. Prefer Model B in TR38.901 |
| CATT | a) This depends on the sub-use case. While it is true that spatial consistency modelling is important for time domain prediction, it is not needed for spatial domain prediction. It can be an optional choice, which is highly recommended for time domain prediction case. |
| LGE | a) Support. It is needed for time domain prediction.  b) Same view as Apple and vivo. Section 7.6.3.1 in 38.901 is enough. |
| Ericsson | 1. Yes, It should also be modeled for spatial beam domain prediction. 2. Our preference is procedure B |
| ZTE, Sanechips | a) It is necessary for spatial consistency modelling to ensure that propagation parameters maintain continuity across multiple realizations. This is particularly important for predicting future optimal beams when the UE moves along a trajectory.  b) We are open to procedure A and B. Per our understanding, procedure A in TR38.901 is more complex than procedure B, yet closer to reality. Compared with procedure B that channel parameters such as delay/power/angles are independent for different UE positions, procedure A applies an iterative algorithm to update the propagation parameters with a restricted moving distance of the UE between consecutive channel realizations. |
| CAICT | 1. Y. If channel prediction is simulated, spatial consistency should be considered. 2. We are open to both procedure A or B. |
| Samsung | a) Yes, spatial consistency model is necessary here.  b) Procedure A for spatial-consistent UE mobility modeling in clause 7.6.3.2 can be used for UE to update the channel cluster power/delay/angles, which helps to generate the consistency for UE with movement. |
| Fujitsu | Spatial consistency should be modeled for time domain beam prediction. It’s not necessary for spatial domain prediction. |
| CMCC | 1. Yes   b) We prefer to use procedure B, since it is closer to realistic scenario. |
| MediaTek | 1. Yes 2. We believe that spatial consistency for mobility needs to be modeled. Procedure B in section 7.6.3.2 of TR38.901 is preferred. |
| Huawei/Hisi | 1. Yes spatial consistency should be modeled. We are open to discuss the details |
| InterDigital | a) Yes as optional. Any company applied a spatial consistency model can report it in their contribution.  b) Procedure B |
| Lenovo | 1. Yes, needed for time domain beam prediction. We think it is not needed for spatial domain beam prediction. 2. We think Procedure A would be sufficient. |
| Qualcomm | 1. Yes, it should also be modeled for spatial (+ temporal) beam prediction 2. Procedure A incorporates temporal evolution of the channel and is our preferred option. We are also open to consider procedure B in which spatial information of Rx positions is taken into account in updating channel realizations which may be beneficial in modeling UE movement on a given trajectory |

### 1.2.2 Trajectory model for UE mobility

Some companies pointed out that impact of discontinuous UE trajectory model needs to be considered.

* Samsung [9]: The impact of discontinuous UE trajectory model on the beam prediction performance shall be considered for UE mobility model selection.
* Nokia/NSB [19]: A suitable UE mobility model is needed for the spatial-temporal domain beam prediction study.
  + As shown in table 2.4-1, the current UE mobility model for spatial-temporal beam prediction assumes a straight trajectory traveled by the UE with a random direction and channel spatial consistency type-A defined in TR 38.901.
* Intel [20]: The UE trajectory should be sampled at least at the minimum decorrelation distance of the large-scale parameters corresponding to the scenario of evaluation.

Moreover, in [9], there options are proposed for the trajectory model for UE mobility:

* Observation 1-1: For the trajectory model for UE mobility, at least the following options exist:
  + Option #1: Linear and fixed trajectory model, e.g., the intra-cell mobility model in Table 2 of R1-2007151.
  + Option #2: Linear trajectory model with random direction change.
  + Option #3: Linear trajectory model with random and smooth direction change.

Based on the above views, the following questions can be discussed for UE trajectory model.

#### FL1 Medium Priority Question 1-4

**Question 1-4:**

1. Whether the UE trajectory model in Table 2 of R1-2007151 is sufficient for the training and/or evaluation of time and/or spatial domain beam prediction?
2. If the answer of a) is no, how to define the trajectory model including the following two options?
   * Option #2: Linear trajectory model with random direction change.
   * Option #3: Linear trajectory model with random and smooth direction change.

|  |  |
| --- | --- |
| Company | Comments |
| Apple | We think it depends on the use case. For spatial domain beam prediction, this is not needed. For time domain beam prediction, it depends on the duration of the monitoring window. If the duration is small, a fixed direction (the direction is randomly generated for each UE) for a UE is enough; if the duration is large, we can consider option 2 or option 3. |
| Nokia, NSB | a) No, it is not sufficient, especially for training and/or evaluation of spatial domain beam prediction.    b) Trajectory model for time beam prediction can be defined considering the distance BS-UE taken from uniform distribution on [dmin, ISD/2], where d\_min is min distance BS-UE, as well as the angle that defines the UE trajectory crossing the sector taken from uniform distribution on [0, 360].   * Spatial beam prediction training and evaluation data requires multiple UEs dropped uniformly in random points of the sector. Then, the UE may adopt the trajectory model defined for time beam prediction. * Option #2 requires an additional parameter that defines the direction Update Distance as well as the turn Probability of the UE after travelling the direction Update Distance. * Option #3 requires in addition of the parameters defined for option#2, an additional parameter that defines turn Angle Limit, to have a smooth changing of UE direction.   Also, we think a clear separation on EVM for training data generation and performance evaluation is needed, and the distance for the trajectories used for training data generation can be short, depending on UE speed. In this case, linear and fixed trajectory of short distance for training data generation shall suffice. Of course, for performance evaluation, the trajectory shall be much longer and more versatile.  Considering this aspect, we can consider another option,  Option#4: Training on random orientation straight-line trajectories. Evaluate on options#1/2/3. |
| vivo | For temporal prediction, ‘No’ to the option1 since it would create overfitting issues.  We would be fine with either Option2 or Option3. UE trajectory can be modeled as in 37.885, and random direction change or smooth direction change described in option 2 and option 3 can be further modeled based on the trajectories defined in 37.855. |
| Intel | Q1-4a: No. The trajectory is relevant only for temporal domain beam prediction which needs collection of mobility trace. The Rel-17 MIMO EVM has a trajectory model with a highway use case which may not be the only relevant case and using this fixed trajectory will lead to overfitting.  Q1-4b: Consider more general UE trajectory modeling where UE is randomly dropped within a deployment, a random direction of travel is chosen and moves in a straight line for a fixed time at a given speed. Many such traces need to be collected to form the dataset for temporal beam prediction. |
| OPPO | a) Yes. But for beam prediction in spatial domain, it is not necessary to model the UE trajectory. |
| AT&T | Agree with Intel and Nokia |
| CATT | Agree with OPPO. |
| LGE | We also think UE trajectory model can be up to each sub use case. For non-HST scenario, either option 2 or option 3 is fine for us. |
| Ericsson | a) For spatial domain we think it could be sufficient.  B) Sharing the view by Vivo that it might create too optimistic results, due to overfitting. Investigating option 2 and 3 should not be precluded. Note that, in general, UE moving at **constant** speed and/or constant rotation should lead to more predictable behavior than dynamic speed/rotation. Another option is to enable modelling of non-constant UE speed. |
| ZTE, Sanechips | We agree with Intel and OPPO that the UE trajectory model is not necessary for spatial-domain beam prediction. For the time-domain beam prediction, we may consider more realistic or complex UE trajectories (such as curve, circular or a combine trajectory) since straight-line trajectory would be rather restricted. |
| CAICT | Y. We believe simple UE trajectory model is enough. |
| Samsung | a) No. For this very simplified UE trajectory model, the randomness is overlooked, and solely based on that, the evaluation will be oversimplified to be justified.  b) we prefer both options2 and 3 at least in the phase of performance evaluation. Both option 2 and 3 are proposed to avoid unrealistic turn over by following a fixed trace, which is not likely to happen. One example of Option-2 can be:   * UE moving trajectory: UE will move straightly along the selected direction to the end of an 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. |
| Fujitsu | It depends on the sub use cases. For spatial domain beam prediction, it’s not necessary to model UE trajectory. |
| CMCC | 1. No. UE trajectory model in Table 2 is not enough flexible. 2. We think it depends on the scenario. For urban scenario, more general UE trajectory model to allow random direction of travel is preferred. For highway/HST use cases, fixed trajectory is sufficient. Both scenarios can be considered. |
| MediaTek | 1. No. At least for time-domain beam prediction, we need to model some kind of UE moving direction change. 2. We would be fine with Option 2, 3, or even other more complicated trajectory modeling. |
| Huawei/Hisi | Not needed for spatial domain BM.  For time domain we agree with other companies that it might depend in the application scenario.   1. For time-domain prediction, it is important to also decide the UE trajectory and how to model spatial consistency. These issues shall not be decoupled for time domain beam prediction. From our point of view, the UE trajectory model from R1-2007151 is not sufficient. For the question b), we are open to both choices but it depends on other simulations assumptions that still remain to be decided. |
| InterDigital | Agree with Intel and Nokia |
| Lenovo | 1. UE trajectory modeling is required only for time domain beam prediction and not for spatial domain beam prediction. 2. Prefer having Option #2 for the UE trajectory modeling. |
| Qualcomm | a) No, due to the fixed nature of the trajectory, the dataset may not be generic enough and the AI/ML model will be subject to overfitting.  b) To avoid the overfitting problem mentioned above, random direction change is desirable, and Option #3 is the most preferred due to providing more flexibility in data generation process and being less prone to overfitting.  In our view, UE orientation should also be taken into account in modeling UE trajectory, as it will affect the beam selection/prediction. Similar to the UE direction in Option #3 which is random and smooth, UE *orientation* can also be modeled to be random and smooth, mimicking realistic real-world scenarios. |

### 1.2.3 Others aspects for data generation or evaluation

The following aspects were discussed and proposed:

* Samsung [9]: For evaluation of beam prediction in the time domain, in addition to the spatially-consistent mobility modelling and UE trajectory modelling, the selection of UE speed, blockage [Section 7.6.4 in TR 38.901], and/or UT rotating [Section 7.6.7 in TR 38.901] should be considered.
* Lenovo [18]: The beam management AI/ML model evaluation methodology should support evalutations with realistic modeling of UE mobility and rotation, spatially consistent channel models, different UE speeds, and single-beam/panel and multi-beam/panel scenarios for evaluating a model's robustness and adaptability.
* NVIDIA [21]: Additional simulation methodology for generating synthetic data, such as digital twins, should be explored for the study of AI/ML based algorithms for beam management.

Based on the above views, the following questions can be discussed for other assumptions:

#### FL1 Low Priority Question 1-5

**Question 1-5:**

1. What other assumptions or parameters need to be defined for SLS based dataset generation and evaluation for beam management?

Note: Mixed scenario(s) for generalization can be discussed in section 2.2.1. This question only focuses on single scenario case for dataset and performance evaluation.

|  |  |
| --- | --- |
| Company | Comments |
| Apple | We suggest adding UE rotation with speed of 360 degree per second |
| Nokia, NSB | UE rotating could be considered as an additional evaluation assumption and should be considered with static UE. |
| vivo | Share same view as apple, UE rotation should be considered for temporal domain beam prediction, and we prefer 50 r/m (round per minute) |
| Intel | OK to consider UE rotation, although 360 degrees per second seems very fast. |
| AT&T | Ok to consider UE rotation |
| CATT | Ok to optionally consider UE rotation but should be in lower priority. |
| LGE | UE rotation can be optionally considered, if needed. |
| Ericsson | We are ok with adding a UE rotation. It should be clarified around what UE axis the rotation is made. Also defining reasonable numbers for such rotation. 360 degree per second sounds very fast. |
| ZTE, Sanechips | Ok to consider UE rotation. It is suggested that the change of UE rotation (if needed) shall follow a simple or fixed pattern, as stated by Apple and vivo. Excessive random rotation during movement would cause the adopted AI model to require huge samples to converge. |
| Samsung | UE rotation, blockage, and random UE speed over time can be considered. We shared the similar view as Intel that 360 degree per second could be too fast. |
| CMCC | Open to consider UE rotation |
| MediaTek | UE rotation may be considered as an option. |
| InterDigital | Fine to consider UE rotation if is optional. |
| Lenovo | UE rotation modelling as in TR 38.901can be considered. |
| Qualcomm | Suggest adding UE rotation for temporal and spatial + temporal beam prediction use cases. |

## 1.3 Dataset generation and evaluation assumptions with LLS

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

* Samsung [9]: Link-level simulation can be used for the evaluation of beam prediction in the time domain.
  + For the evaluation of beam prediction in the time domain, the CDL extension used in the Rel-17 HST-SFN evaluation should be a good starting point in link-level evaluation.
* Samsung [9]: Consider Table A.2.5-2 (for SLS) and Table A.1.6.4 (for LLS) of TR 38.802 as the starting point for Rel-18 AI/ML for beam management evaluation, based on which necessary modifications/simplifications can be further discussed for the specific evaluation purpose and/or sub use case.
* Intel [20]: For spatial domain beam prediction, CDL channel models should be used for dataset generation and RSRP can be used as the beam specific parameter for the dataset.
* Intel [20]: Assumptions on gNB and UE antenna arrays and beamforming should be aligned across companies for common dataset generation
* Intel [20]: Dataset normalization based on specific sub-use case should be reported by companies

Based on the above views, the following questions can be discussed for LLS for AI/ML-based beam management.

#### FL1 Medium Priority Question 1-6

**Question 1-6:**

1. Whether LLS can be used as a complementary evaluation methodology for AI/ML in beam management?
2. If the answer is yes, whether Table A.1.6.4 (for LLS) of TR 38.802 can be used as the starting point for Rel-18 AI/ML for beam management evaluation?

|  |  |  |
| --- | --- | --- |
| Company | Y/N | Comments |
| Apple | N | At least SLS should be the baseline. LLS may cause overfitting problem. |
| Nokia | N | Disagree that LLS are needed for this SI. The BM use case requires mainly beam measurements data (L1-RSRP) for multiple beams and UEs in different locations of the sector. We don’t see the advantage of using LLS for this use case since LLS will simulate the BS-UE transmission link with too many details (down at the bit-level), which is too much simulation burden for the BM study, and it can be avoided. On the other hand, we believe that SLS enables us to generate a large amount of data for UEs in different positions for a single network drop and enables us to diversify the data to improve the ML performance during training while enabling to test the ML model generalization capabilities during the evaluation phase. |
| vivo |  | We are open to discuss |
| Intel | Y | LLS channel models i.e., CDL models can be used to generate RSRP data for spatial domain beam prediction. LLS for performance evaluation may not be needed |
| NVIDIA |  | Interested companies can present the results with sufficient level of description. |
| OPPO |  | SLS is the baseline. LLS seems not necessary. Anyway, companies can bring up LLS evaluation results if they are willing to. |
| CATT | N | We don’t need to duplicate the evaluation effort since SLS evaluation is sufficient. |
| LGE | N | Prefer to have SLS as the baseline. |
| Ericsson | N | No need to address LLS. |
| ZTE, Sanechips |  | SLS should be the baseline and LLS can be optional provided for preliminary evaluation. |
| CAICT | N | SLS is preferred. |
| Samsung | Y | We share an opposite view with Nokia. In our discussion paper, we proposed reusing the CDL-extension used in Rel-17 HST-SFN to get L1-RSRP data for multiple beams from different UE locations considering UE trajectory. **The advantage is that this LLS does not require the spatial consistent modeling described in TR 38.901, so we can reduce our workload for simulation.** Also, we can avoid the overfitting issue by adopting random trajectory modeling, random UE’s boresight, random UE speeds and so on. Moreover, if we study and discuss the beam management for high speed users including HST, highway in the future, this LLS modeling would be useful. Thus, reusing the existing LLS can be adopted as a simpler metric for evaluation at least for Phase 1 evaluation (e.g., to evaluate Top-N accuracy, RSRP gap, etc.), while FFS Phase 2 evaluation (to evaluate UPT, etc.). |
| Fujitsu | Y | 1. Yes 2. Yes |
| CMCC |  | Open to use LLS for spatial domain beam prediction. |
| MediaTek | N | Beam management is to deal with UE mobility, which is usually studied using SLS. |
| HW/HiSi | Y | 1. LLS shall be considered. Especially for spatial domain BM 2. The table can be used as a starting point |
| InterDigital |  | Do not support LLS as an evaluation methodology as LLS does not accurately show the actual impact of beam prediction. |
| Lenovo | Y | 1. Yes, at least for the beam prediction evaluation. 2. Yes |
| Qualcomm | Y | We are open to discussing the prospect of optional LLS but prefer SLS since the channel models cover a broader range of scenarios. With LLS there’s a concern that the dataset may not be generic enough, and therefore there’s a risk of overfitting. |

## 1.4 AI/ML model related aspects

Some companies suggested to define reference AI/ML model for calibration, while some companies believed no need to introduce a reference AI/ML model:

* Huawei/HiSi [1]: The calibration of the AI/ML model and dataset for beam management evaluations can be achieved by aligning simulation assumptions and calibrating intermediate results such as beam prediction accuracy.
* Ericsson [4]: No need to introduce a reference AI/ML model for calibration
* Vivo [7]: Fully-connected neural network with a limited number of model parameters for beam management case is considered as the basic model. Other models can also be considered if found beneficial.
* Apple [15]: During the use case study phase, it is not necessary to define a common AI architecture.
* NVIDIA [21]: Baseline AI model(s) should be identified for the purpose of calibration in the study of AI/ML based algorithms for beam management.
* Fujitsu [24]: Calibrate the performance of the AI/ML method, a common reference model is recommended, and its generation procedure needs to be studied.

Based on the above discussion, it is better to discuss whether a reference AI/ML model needs to be defined.

#### FL1 Medium Priority Question 1-7

**Question 1-7:**

1. Whether a reference AI/ML model needs to be defined, and why?
2. If the answer is yes, please explain the purpose to define the reference AI/ML model.

|  |  |
| --- | --- |
| Company | Comments |
| Apple | a) At current stage, no. |
| Nokia, NSB | 1. No, only AI/ML model inputs/output need to be defined but ML model details will be up to companies' implementation. |
| Xiaomi | No. companies can report the NN architecture, the input/output of the AI/ML model. |
| vivo | For cross check purposes, we should support reference AI/ML model for calibration or companies should open the details as much as possible. |
| Intel | At this stage, no. Meaningful non-AI/ML baselines should be considered for fair comparison |
| NVIDIA | Yes, a baseline model would facilitate calibration. |
| OPPO | We are open to it. However, it may be difficult to define a reference AI/ML model. Companies are encouraged to share their models used in the evaluation for cross check and calibration. |
| CATT | We are open to consider it, at least for the purpose of calibration. |
| LGE | No, companies can provide there assumption on AI/ML model. |
| Ericsson | Not needed in case the model training methodology and model description are properly defined |
| ZTE, Sanechips | Rather than defining a reference AI/ML model, more importantly each company may provide a general and detailed description of the AI/ML models adopted for cross-validation. |
| CAICT | It might be hard to directly define common reference model. It is proposed that AI/ML model structure description is provided along with simulation results for double check. |
| Samsung | No. we don’t see the need, at least in this stage. Especially in this stage, due to the use case also being studied, it is hard to provide and justify a certain reference AI/ML model is suitable. |
| Fujitsu | Yes. The purpose to define the reference AI/ML model is to easily calibrate the simulation results |
| CMCC | We support to define a reference AI/ML model for calibration. If no, we have no idea how to carry out calibration. |
| MediaTek | 1. At this stage, no. |
| HW/HiSi | No, at least not in the starting phase. Companies could provide a general description of their model. |
| InterDigital | No. |
| Lenovo | A reference AI/ML model can be considered for the calibration purpose. |
| Qualcomm | a) No, generating data based on TR 38.901 and an agreed set of EVM assumptions is sufficient. There should be no need for further calibration of the data. Also, there is no need to align the AI model used by companies. |

On the other hand, the following aspects related to AI/ML models and input/output of AI/ML models were proposed:

* Huawei/HiSi [1]: For AI/ML-based spatial domain beam prediction evaluation, adopt the RSRP for sparse beams and the optimal beam ID as the training inputs.
* Ericsson [4]: Evaluation should provide a description of ML/AI model, model input/output, along with the training methodology
* Samsung [9]: For AI/ML training and inferring, the detailed AI/ML model, and loss/reward function design should be decided by companies, while the relevant information should be explained by companies:
  + AI/ML model selection: e.g., DNN, CNN, LSTM with the number of layers, and the number of nodes in each layer, and some necessary description for the node connection, etc.
  + Input of AI/ML model.
  + Output of AI/ML model.
* LGE [13]: Consider following information as input data for AI/ML based DL Tx beam prediction in time domain.
  + For gNB side AI/ML, gNB measurement (e.g., UE position/trajectory, UL measurement) and UE beam related reporting.
  + For UE side AI/ML, UE own measurement (e.g., UE position/velocity, UE orientation/rotation, DL beam/CSI measurement).
* Lenovo [18]: Any additional side information (e.g., measurements from non-3GPP technologies) needed for a proposed beam management AI/ML model should be stated.
* Ericsson [4]: Model validity area should be described, if the AI/ML model is trained for a cell (cell-specific), multiple-cells, or for the entire scenario (valid for all cells).
* Ericsson [4]: Models are trained offline (no sequential update of the model)

Besides, the following training methodology were proposed:

* Ericsson [4]: Training methodology could for example include the training optimization function, loss function and/or number of training/test samples
* Vivo [7]: Training dataset and validation dataset should be collected from different drops to keep data independence.

#### FL1 Medium Priority Question 1-8

**Proposal 1-8:**

* **Each company should provide a description of ML/AI model, [model input/output], along with the training methodology.** 
  + **FFS on the details**

**Question 1-8:**

1. Whether proposal 1-8 can be adopted? Why?
2. What parameter(s)/aspect(s) of AI/ML model(s) need to be reported by each company?
3. Whether to define model input(s)/output(s) subjected to each sub-use case or the model input(s)/output(s) of each sub-use case can be reported by each company for AI/ML in BM?
4. What aspects need to be reported or defined for training methodology?

|  |  |  |
| --- | --- | --- |
| Company | Y/N | Comments |
| Apple | Y | Companies can report the NN architecture |
| Nokia, NSB | Y | a) Agree to specify the ML model input and output, since the exchange of the input and output data may have a standard impact, especially when UE model inference is done at the UE side. The ML model complexity may also be reported by each company.  b) Description of ML/AI model input/output including extra information (e.g., measurements not considered in 3GPP) needed for ML model input, model validity (single sector/multi sectors).  c) ML model input may differentiate between subcases and therefore should be specified.  d) Dataset size (e.g. number of UEs used for training/validation/testing), model trained for single sector or multi sectors. |
| Xiaomi | Y | For b), Companies can report the NN architecture, and the parameters such as number of layer etc..  For c), Also the model input/output need to be reported by each company and need to be specified if have a standard impact.  For d), For training methodology, for example, dataset size and model trained for single cell or all cells can be reported. |
| vivo | Y | 1. Support, for calibration purpose. 2. The detail of AI Model architecture should be reported, such as model type, layer number, quantization and loss function. 3. Parameter type and corresponding number of model input/output should be reported.   Data independence should be defined in dataset construction between training data and validation data. |
| Intel | Y | Q1-8.a: Yes  Q1-8.b: In addition to description of AI/ML model, input/output, companies should also report dataset normalization  Q1-8.c: Yes input/output and model description should be provided for each sub-use-case |
| NVIDIA | Y | Yes, description of AI/ML model should be provided. Details can be FFS. |
| OPPO | Y | a) Support in principle.  b) At least the high-level description for AI/ML model should be reported. Detailed information is up to each company. Companies are encouraged to disclose more details.  c) Model input/output should be reported.  d) It is up to each company. |
| CATT | Y | a) Yes  b) The input(s)/output(s) of AI/ML model for each sub-use case can be reported by each company.  c) The output may be specified. While at the first stage, we are open to see different input(s) reported by companies.  d) At least the details of training data set should be reported, i.e. channel model, dataset size. |
| LGE | Y | a) Yes.  b) NN architecture can be provided.  c) Either way is fine. However, there is similar ongoing discussion in AI 9.1.3.2. |
| Ericsson | Y | 1. Agree. This allows for some degree of cross checking of results among companies and the possibility to assess the complexity, memory consumption, FLOPS etc. 2. Model description (type of model, parameters), Model input description, Model output description. The method used in e.g. IEEE papers on ML can be a guideline. 3. Agree – the input(s)/output(s) should depend on each subuse case 4. Loss function, optimization function, Dataset description (Training/Test data), Number of training/test samples, Model validity area |
| ZTE, Sanechips | Y | Parameters to be reported by each company at least include model input, model output, dataset size, and other general description of the employed AI/ML model. |
| CAICT | Y | The detail of AI model should be mainly focus on the model structure. |
| Samsung | Y | a) Yes for verification of each companies AI-model.  b) at least, the structure of AI/ML model (including type of neural network, number of layers, number of nodes in each layer, connections, etc), input/output information should be reported. Providing details such as post/pre-processing, training strategy, and so on can be up to each company.  c) model input/output should be reported, to study specification impact.  d) generation method of training data set should be provided from each company. Also, model scalability of each company should be reported. Furthermore, the loss function should also be encouraged to be reported for companies’ training. |
| Fujitsu | Y | Companies need to report AI/ML model structure, input/output, and training methodology |
| CMCC | Y | a) Yes  b) Besides input/output, companies are encouraged to report the NN architecture and network parameters.  c) Even for one sub use case, the input/output may different, thus it can be reported by company. |
| MediaTek | Y | 1. Companies can report the NN architecture (which can be generated by AI/ML tools) 2. Input/output format should be reported   for training methodology, companies may report data set size and the scope of trained model (generic, cell-specific, UE-specific, etc.)   1. Companies are encouraged to provide more detailed parameters for training such as batch size, learning rate, optimization technique, etc |
| HW/HiSi | Y | 1. Yes 2. Requires more discussion 3. Model input(s)/output(s) of each sub-use case can be reported by each company for AI/ML |
| InterDigital | Y | 1. High level model description 2. Input(s)/output(s) of each sub-use case can be reported by each company 3. This can be optional and up to each company. |
| Lenovo | Y | 1. Yes. Having the details of the AI/ML model helps, understanding the merits and merits and demerits of the proposed AI/ML method and the generic applicability of different ML methods for the BM use case. 2. We consider that following Model Characterization Card (MCC) (presented in [R1-2204416](https://www.3gpp.org/ftp/TSG_RAN/WG1_RL1/TSGR1_109-e/Docs/R1-2204416.zip)), would be an effective way of capturing different aspects of an AI/ML method. More details about the model (such as NN structure (e.g., CNN or RNN), input/output, hyper-parameters etc.) may be reported as additional details.   **Proposed AI Model Characterization Card**   |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | | **AI/ML Proposal** | **Model Training/Validation Phase** | | | | **Model Deployment /Inference Phase** | | **Model Monitoring/Update Phase** | | | | **Model Training** | **Learning Paradigm** | **Data Set** | **Data Collection** | **Deployment** | **Inference Data collection** | **Model Monitoring** | **Model update paradigm** | **Update Data collection** | | **The Name of the proposed scheme** | Offline (simulation) | Centralized | Simulated data | Only simulated data | In network | Legacy signaling - no spec impact | No monitoring needed | Centralized | No field data needed | | Pre-trained (simulation) + fine-tuned (field) | Federated | Offline field data | Legacy signaling - no spec impact | In UE | Collaboration b/w a UE & gNB | At corresponding nodes | Federated | Offline field data | | Field | Distributed | Online field data | Collaboration b/w a UE & gNB | Split b/w network and UE | Across multiple nodes | At another node | Distributed | Online field data | |  |  |  | Across multiple nodes |  |  | Initiated periodically |  |  |  1. Yes, the model inputs and outputs need to be specified for the use case of BM 2. Details of Training paradigm: Such as offline or online, or initial offline training plus online finetuning, Centralized/Federated/Distributed training etc. |
| Qualcomm | Y | a) Yes, in principle, with the following description  b) A high-level description of AI/ML model can be provided. Providing more detailed description should be voluntary.  c) The input and output to the AI/ML model can be voluntarily provided by companies. The pre- and post-processing for AI/ML model input and output can be up to implementation. In our view, reporting the KPIs is sufficient, and the explicit input and output of AI/ML model can be up to companies to provide.  d) Can be voluntarily provided by each company. Candidates: dataset size, training on a given set of UEs, testing on another set, whether model is trained for single sector or multiple sectors |

## Others

### 1.5.1 HST in FR2

Two companies mentioned beam management enhancement for high-speed train in FR 2.

* ZTE [2]: Consider predictable mobility for beam management as an enhancement aspect for improving UE experience in FR2 high mobility scenario (e.g., high-speed train and high-way).
  + Study and evaluate the feasibility and potential system level gain on predictable mobility for beam management based on the identified scenario(s).
* PML [3]: Consider predictable mobility for beam management as an enhancement aspect for improving UE experience in FR2 high mobility scenario (e.g., high-speed railway and high-way) in a Rel-18 WI.
  + Study the implementation and design of predictable mobility for beam management in various scenarios.
  + Evaluate the performance gain and cost of predictable mobility for beam management in a more concrete and comprehensive manner.

Based on the above views, the following questions can be discussed for high-speed train scenario in FR 2 for AI/ML for beam management.

#### FL1 Medium Priority Question 1-9

**Question 1-9:**

1. Whether HST can be considered as one of the scenarios for AI/ML for beam management? If the answer is yes, which shall be the baseline tool for dataset generation and performance evaluation, e.g., LLS or SLS?

|  |  |  |
| --- | --- | --- |
| Company | Y/N | Comments |
| Apple | N | We think current scenario is enough. At least HST should not be the baseline scenario. |
| Nokia, NSB | N | a) No need to include HST scenario in the set of simulations assumptions. We believe that our focus should concentrate on the Dense Urban scenario, which is supported by most companies |
| vivo |  | Open to discuss |
| Intel | N | Agree with Nokia |
| NVIDIA |  | Interested companies can present the results with sufficient level of description. |
| OPPO |  | Low priority |
| CATT | N | No need to include HST scenario. |
| LGE | N | Similar view as Nokia. |
| Ericsson | N | No need to study HST at this stage. |
| ZTE, Sanechips | Y | HST can be considered as one of the scenarios for AI/ML for beam management. As analyzed in our companion contribution R1-2203250, the high speed mobility and short beam dwelling time in HST lead to huge RS overhead for beam tracking and large latency for beam indication, which makes it imperative to introduce AI/ML technology for time-domain beam prediction. Besides, according to our preliminary simulation results, RS overhead and latency can be significantly reduced with AI/ML based beam prediction. Meanwhile, cell-edge UE UPT and mean UPT are also improved obviously. Regarding the dataset generation and performance evaluation, SLS can be adopted as the baseline tool.  Additionally, the scenario related problem may be discussed together in ‘FL1 High Priority Question 1-2’ instead of discussing it individually. |
| CAICT |  | Low priority but open to discuss. |
| Samsung |  | We are open to discuss this scenario. But general scenarios with low and medium UE speeds should be prioritized. |
| Fujitsu | N | To limit the workload, HST scenario is not recommended in initial stage. |
| CMCC | Y | We think HST can be considered as one of the scenarios. The baseline tool should be aligned with time domain beam prediction. |
| MediaTek | N | Dense Urban scenario should be our focus. |
| HW/HiSi | N | We think that scenarios with lower UE speed have higher priority. |
| InterDigital | N | No need to include HST scenario. |
| Lenovo | N | HST is a very special scenario of BM and need not be given a separate consideration. |
| Qualcomm | N | a) No need to consider HST |

### 1.5.2 Other consideration for evaluations

Several companies propose to define two phases for evaluations:

* CATT [6]: For AI/ML based beam management, the following evaluation can be considered:
  + Intermediate evaluation: Evaluate the result of beam selection from the AI model, or the overhead reduction due to AI/ML-based approach;
  + Final evaluation: Evaluate the system-level transmission performance based on the output of AI model.
* Vivo [7] Intermediate results for performance comparison across companies can be considered.
* Samsung [9]: The following two-stage approach is adopted for gNB/UE beambook design:
  + Stage-1: Alignment on the number of beams (as input/output for AI/ML beam prediction)
  + Based on the initial evaluation results, and to be concluded by RAN1 #110.
  + Stage-2: Based on the agreed number of beams (as input/output for AI/ML beam prediction), companies provide the simulation results:
    - Companies shall disclose the detailed beambook design: beam directions/beamwidth.
    - Companies shall disclose the detailed beam measurement and report configuration.

This can be discussed together with performance KPIs.

# 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.

## Performance KPIs

### 2.1.1 Beam measurement related KPIs

Beam measurement or selection accuracy were proposed by most of the companies, including:

* ZTE [2]: Top-K candidate beams with higher predicted RSRP can be filtered out for refined small-range beam sweeping, resulting in a relatively good trade-off between training overhead and performance.
* ZTE [2]: Beam prediction accuracy can be used as the performance indicators at the early stage, which may include top-1/top-K beam prediction accuracy, average RSRP difference, and CDFs of RSRP difference between the AI-predicted beam and ideal beam.
* Ericsson [4]: Use case performance metrics could include metrics defined in Rel-17 EVM for BM including a potential energy saving metric, for example using the models defined in [4].
* InterDigital [5]: The following parameters should be considered to accurately evaluate benefits of AI/ML aided beam management.
  + Difference between estimated qualities and actual qualities
  + Optimal beam selection accuracy (%)
  + System performance based on the selected optimal beams
* CATT [6]: Intermediate KPI: Probability of identifying the best beam (pair), average power loss between the selected beam (pair) and the optimal beam (pair), and the ratio of overhead reduction;
* Vivo [7]: The 4 proposed RSRP-related KPIs should be considered for performance evaluation and comparison among different selected algorithms.
  + RSRP difference 1 for all predicted beam
  + RSRP difference 2 for predicted beam set
  + beam pair prediction deterioration
  + beam pair prediction accuracy
* xiaomi [8]: KPI:
  + KPI#1: The probability of right best Tx beam predicted by AI.
  + KPI#2: The L1-RSRP gap between the predicted best Tx beam and the real best Tx beam.
* Samsung [9]: For the performance-related KPI for AI/ML based beam prediction (in both time and spatial domains), the following two KPIs can be utilized, especially for Phase-I:
  + Accuracy of predicted best beam within the Top-N from genie-aided beam measurement.
  + Performance gap from the genie-aided beam measurement.
* Samsung [9]: In this SI, study and evaluate the performance of AI/ML beam prediction in the time domain in terms of the top N beam prediction accuracy and the overhead and latency reduction.
* OPPO [10]: Adopt L1-RSRP as performance metric for AI/ML beam prediction as a starting point.
* OPPO [10]: Study and evaluate the correct beam prediction rate for AI/ML beam prediction in spatial and time domain as one of the key performance metrics.
* Beijing Jiaotong University [11]: Beam selection accuracy, mean square error (MSE), throughput and system capacity, need to be considered as the KPI for the evaluation of beam selection.
* Beijing Jiaotong University [11]: Overhead, latency and beam selection accuracy of existing L1-RSRP measurement scheme, need to be considered as the benchmark of AI-based beam selection.
* Futurewei [12]: Include Top-1 prediction accuracy (Top1\_acc) as one of the evaluation metrics for AI/ML-based beam management use case.
* Futurewei [12]: Include Top-K (e.g., Top-3) prediction accuracy as one of the evaluation metrics for AI/ML-based beam management use case.
* Futurewei[12]: Include measured RSRP as one of the evaluation metrics for AI/ML-based beam management use case.
* CAICT [14]: The success rate of selecting best link quality beam pair(s) under a given number of measurements is used for performance comparison between AI/ML based algorithm and non-AI/ML based algorithm.
* CAICT [14]: The performance of the optimal beam pair(s) prediction by AI/ML model(s) could also be considered.
* CAICT [14]: Inference accuracy should be the main KPI for AI/ML based beam management algorithm.
* Apple [15]: The KPI for AI based beam prediction could be the beam prediction accuracy and the L1-RSRP distribution for the AI predicted beam.
* Lenovo [18]: KPIs for the AI/ML beam management performance evaluation should include beam alignment success rate, effectiveness (in terms of improvement in L1-RSRP or L1-SINR relative to legacy schemes), and beam failure rate.
* Nokia/NSB [19]: RAN1 shall consider the following KPIs for the ML-based beam management use case,
  + Beam prediction accuracy (Top-1 Prediction, Top-K Prediction, Top-1 and Top-K Prediction with 1 dB margin)
  + RSRP error/difference
* Intel [20]: For beam management use cases, hard metric KPIs like accuracy of best beam index or top K beam index prediction can be considered. Additionally, RSRP of predicted beams should also be considered as a key KPI for performance evaluation.
* NVDIA [21]: Define a KPI which measures beam management accuracy with and without AI/ML based algorithms.
* Qualcomm [23]: The KPIs for temporal beam prediction can be trade-off between reference signal overhead reduction (and/or UE power consumption) and a measure for beam prediction accuracy (such as RSRP prediction accuracy and/or top-M beam selection accuracy) and/or overall system performance (e.g., throughput).
* Qualcomm [23]: The KPIs for spatial (+time) domain beam prediction can be trade-off between reference signal overhead reduction (and/or UE power consumption) and a measure for beam prediction accuracy (such as L1-RSRP prediction accuracy and/or top-M beam selection accuracy) and/or overall system performance (e.g., throughput).
* Fujitsu [24]: The following KPIs are recommended for performance evaluation of beam prediction.
  + Probability of correct prediction
  + Average L1-RSRP difference
* MediaTek [25]: Inter-cell beam management (ICBM) can be considered.
  + KPI can be the accuracy of Cell ID(s) which the predicted best-N beam(s) belong to.
  + Note: ICBM does not necessarily mean mobility. ICBM introduced in Rel-17 feMIMO WI actually does not support mobility (i.e., no serving cell change). Therefore, ICBM should be considered in the scope of AI for BM.

Based on the proposals, the following proposal can be considered.

#### FL1 High Priority Question 2-1

**Proposal 2-1:**

* The following beam prediction accuracy related KPIs for AI/ML in BM can be considered as a starting point:
  + L1-RSRP related KPIs:
    - Top-1/Top-K predicted beams:
      * Average L1-RSRP difference
      * CDF of L1-RSRP difference
      * CDF of L1-RSRP
  + Beam selection accuracy (%) without margin or with 1dB margin.
    - Top-1
    - Top-N
  + Note: Top-K beams are the Top-N from genie-aided beam measurement

**Question 2-1:**

1. Whether proposal 2-1 can be adopted? If no, what else is necessary to be considered and why?
2. Which KPI(s) are preferred as basic KPI(s)?
3. Which KPI(s) can be optional reported by each company?

|  |  |  |
| --- | --- | --- |
| Company | Y/N | Comments |
| Apple | Y |  |
| Nokia | Y | a) Support.  b) Average L1-RSRP difference, Top-1 beam selection accuracy  c) CDF of L1-RSRP difference, CDF of L1-RSRP, Top-N beam selection accuracy |
| Xiaomi | Y | Support a)  For b), Average L1-RSRP difference and Top-1/K beam selection accuracy.  For c), CDF of L1-RSRP difference and CDF of L1-RSRP. |
| vivo | Y | 1. Support 2. For RSRP related-KPI, we prefer average L1-RSRP difference and CDF of L1-RSRP difference as basic KPI, and both Top-1 and Top-N beam selection accuracy without margin or with 1dB margin can be considered as basic KPI.   For RSRP difference issue, at least three types of RSRP difference can be defined:  Type 1: Predicted RSRPs of top-k beams in predicted set – actual RSRPs in labelled set with the same K-th beam IDs;  Type 2: Predicted RSRPs of top-k beams in predicted set – actual RSRP of best beam in testing set;  Type 3: Actual RSRPs in labelled set of Top-k beams ID selected from predicted set -actual RSRP of best beam in testing set;  Thus, we believe the details of RSRP difference should be discussed further. |
| Intel | Y | a. Support the proposal  b. Top-1/N accuracy and Average L1-RSRP difference can be baseline  c. CDF of L1-RSRP difference can be optionally reported |
| NVIDIA | Y | a) Support  b) CDF of L1-RSRP difference, Top-K beam selection accuracy |
| OPPO | Y | a) Support  b) Average L1-RSRP difference, CDF of L1-RSRP difference, Top-1 beam prediction accuracy  c) any other KPIs |
| CATT | Y | a) Support the proposal.  b) Average L1-RSRP difference of the Top-K beams can be set as the basic KPI.  c) All other KPIs can be optionally reported by companies. |
| LGE | Y | a) Support.  b) Average L1-RSRP difference and Top-1/K beam selection accuracy.  c) CDF of L1-RSRP difference and/or CDF of L1-RSRP. |
| Ericsson | Y | 1. Agree 2. Average L1-RSRP difference and Top-1/K beam selection accuracy. |
| ZTE, Sanechips | Y | 1. Support. 2. Average L1-RSRP difference, CDF of L1-RSRP difference, and Top-1/K beam selection accuracy can be baseline. 3. CDF of L1-RSRP can be optionally reported. Besides, does the L1-RSRP here corresponding to the beam measurement results of different UEs or that of the same UE at different times? |
| CAICT | Y | a) Support  b) Average L1-RSRP difference, CDF of L1-RSRP difference, Top-1 beam prediction accuracy |
| Samsung | Y | a) support  b) Average L1-RSRP difference, Top-N beam prediction accuracy (N=1, and 3 can be selected firstly).  c) CDF of L1-RSRP difference: optionally reported because it is hard to align and compare among companies, and only provide additional information. |
| Fujitsu | Y | Basic KPIs:   * Average L1-RSRP difference and CDF of L1-RSRP difference * Top-1/N beam selection accuracy without margin |
| CMCC | Y | a) Yes  b) Average L1-RSRP difference, Top-N beam selection accuracy  c) CDF of L1-RSRP difference |
| MediaTek | Y | b) Average L1-RSRP difference and Top-1/N beam selection accuracy can be baseline  c) CDF of L1-RSRP and cell selection accuracy for inter-cell beam management case |
| HW/HiSi | [Y] | It should be clarified that the average L1-RSRP difference can also be calculated from difference beams sets. For example if the gain of narrow beams over wide beams should be evaluated. Or if the gains from super resolution non-orthogonal beams vs DFT beams should be evaluated  To capture this, we prefer to add note to the proposal how to understand the calculation of the L1-RSRP difference.  **Proposal 2-1:**   * The following beam prediction accuracy related KPIs for AI/ML in BM can be considered as a starting point:   + L1-RSRP related KPIs:     - Top-1/Top-K predicted beams:       * Average L1-RSRP difference         + Note: l1-RSRP difference can be obtained across different sets of beams       * CDF of L1-RSRP difference       * CDF of L1-RSRP   + Beam selection accuracy (%) without margin or with 1dB margin.     - Top-1     - Top-N   + Note: Top-K beams are the Top-N from genie-aided beam measurement |
| InterDigital | N | a) We prefer to discuss system performance and L1-RSRP together. We propose to include system performance in this proposal.  b) Average L1-RSRP difference, Top-N beam selection accuracy  c) CDF of L1-RSRP difference |
| Lenovo |  | 1. Yes. 2. We think it is important to consider Latency incurred (i.e., time required for Tx-Rx beam selection) as a KPI for evaluating the performance of a proposed BM method and we discussed this in detail in our response to Question 2-3. For example, a scheme may achieve very high beam prediction accuracy, but it may need a long time in finding the beam-pair, rendering it practically useless.   Further, we suggest considering “Beam Failure Rate” as a KPI to evaluate the effectiveness of the AI/ML method’s beam tracking ability (if the AI/ML method also addresses the beam tracking). If the AI/ML model's beam tracking is not good, there would be more instances of beam failure. It can be computed as the number of beam failure events over a long time-period.  All the KPIs need to be reported across a range of SNR/SINR values covering low, moderate and high SNR/SINR regimes. |
| Qualcomm | Y | a) Support  b) top-1/top-N beam selection accuracy and average L1-RSRP difference  c) CDF of L1-RSRP and CDF of L1-RSRP difference |

### 2.1.2 System performance related KPIs

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

* Huawei/HiSi [1] Consider a super-narrow codebook (e.g., with 256 Tx beams) for the evaluation of the AI/ML-based spatial domain beam prediction to improve the system performance (e.g., coverage, throughput) from the network perspective without increasing the maximum number of CSI-RS resources.
* Ericsson [4]: Use case performance metrics could include metrics defined in Rel-17 EVM for BM including a potential energy saving metric, for example using the models defined in [4].
* Vivo [7]: Throughput performance for space/time domain beam prediction are also considered.
* Samsung [9]: For Phase-II, user-perceived throughput (UPT) can be utilized as the KPI to evaluate the performance benefits obtained by AI/ML-based beam measurement.
* Samsung [9]: Shannon capability based simplified model for UPT can be employed in the EVM of AI/ML-based beam management.
* Beijing Jiaotong University [11]: Beam selection accuracy, mean square error (MSE), throughput and system capacity, need to be considered as the KPI for the evaluation of beam selection.
* Nokia/NSB [19]: RAN1 shall consider the following KPIs for the ML-based beam management use case,
  + Beam management measurement overhead
  + Cell throughput (average, 5%ile, 50%ile)
* Qualcomm [23]: The KPIs for temporal beam prediction can be trade-off between reference signal overhead reduction (and/or UE power consumption) and a measure for beam prediction accuracy (such as RSRP prediction accuracy and/or top-M beam selection accuracy) and/or overall system performance (e.g., throughput).
* Qualcomm [23]: The KPIs for spatial (+time) domain beam prediction can be trade-off between reference signal overhead reduction (and/or UE power consumption) and a measure for beam prediction accuracy (such as L1-RSRP prediction accuracy and/or top-M beam selection accuracy) and/or overall system performance (e.g., throughput).

Based on the proposals, the following proposal can be considered.

#### FL1 High Priority Question 2-2

**Proposal 2-2:**

* **System performance is considered as one of the general KPIs for AI/ML in beam management. FFS on the following details:**
  + **UE throughput: CDF of UE throughput, avg. and 5%ile UE throughput**

**Question 2-2:**

1. Whether proposal 2-2 can be adopted? If not, why?
2. Which KPI(s) are preferred as basic KPI(s)? Are they common for all the sub-use cases or subject to some of sub-use case(s)?
3. Which KPI(s) can be optional reported by each company?

|  |  |  |
| --- | --- | --- |
| Company | Y/N | Comments |
| Apple | N | We are open to consider throughput as an optional metric. Usually L1-RSRP should be sufficient for BM related evaluation. |
| Nokia | Y | a) Yes,  b) average cell throughput, 5%-percentile and 50%-percentile UE throughput  c) Packet delay (for the sub-use case of spatial domain beam prediction for throughput and latency reduction) |
| Xiaomi | N | We prefer to consider throughput as an optional metric and we share same view as Apple. |
| vivo |  | Support at least in SI-phase 2 stage for further performance evaluation. |
| Intel | N | The KPIs related to beam measurement should be the baseline and are enough to evaluate AI/ML model performance. System performance KPIs should be optional if at all needed. |
| NVIDIA | Y | a) Yes,  b) CDF of UE throughput, avg. and 5%ile UE throughput |
| OPPO |  | It can be an optional KPI  Average cell throughput, 5%-percentile UE throughput |
| AT&T | Y | 1. Prefer CDF of UE throughput and 5% UE throughput 2. Packet delay |
| CATT | Y | a) Yes  b) CDF of UE throughput shall be the basic one. |
| LGE | N | Prefer to have throughput as optional metric. |
| Ericsson | Y | a) Agree  b) CDF of UE throughput, avg. and 5%ile UE throughput  c) CDF of UE beam failures, Note that time-based beam predictions could be used to mitigate such occurrences |
| ZTE, Sanechips |  | Throughput related metrics such as 5%-edge, 95%-ile, and mean UPT can be evaluated for time-domain beam prediction. |
| CAICT | N | It may depend on the operation mode for AI/ML based BM. We think Proposal 2-1 should be baseline. |
| Samsung |  | We think throughput need to be considered in the end. But for phase I study, this is not necessary. |
| Fujitsu | N | In initial stage, the UE throughput is not necessary since KPIs of proposal 2-1 are sufficient to show the gains of AI/ML-based method. |
| CMCC |  | We prefer to consider throughput as an optional metric. |
| MediaTek | N | L1-RSRP should be sufficient for BM evaluation. UE throughput can be considered as an optional metric. |
| HW/HiSi |  | We are open to discuss system performance |
| InterDigital | Y | As we commented, L1-RSRP is not enough for BM evaluation. We strongly prefer to discuss system performance together with L1-RSRP in a same proposal. |
| Lenovo | N | 1. We think the resulting L1-RSRP/L1-SINR after beam alignment (or, beam selection/pairing) should serve as a good and sufficient metric. However, we are open to having throughput as an optional metric. 2. Overhead and latency, along with beam prediction accuracy metrics of Proposal 2-1, should be considered as basic KPIs. |
| Qualcomm | Y | a) Yes,  b) CDF of UE throughput  Beam selection accuracy and L1-RSRP difference do not capture the full picture and benefit of AI/ML on their own, as they are not representative of the overhead needed to achieve the respective performance, and therefore may lead to misinterpretation of evaluation results. |

### 2.1.3 Other performance KPIs

Overhead reduction and latency were proposed as one of the KPIs for AI/ML in beam management:

* Huawei/HiSi [1]: Consider the evaluation of the AI/ML based spatial domain beam prediction with the target to obtain accurate narrow beams for CSI-RS beam sweeping to reduce the UE’s overhead and power consumption.
* ZTE [2]: AI/ML based solutions are expected to be studied and evaluated to do beam prediction so as to reduce beam tracking latency and RS overhead in high mobility scenarios.
* Ericsson [4]: The study item should investigate how beam predictions can improve robustness and reduce overhead associated to beam management
* CATT [6]: Intermediate KPI: Probability of identifying the best beam (pair), average power loss between the selected beam (pair) and the optimal beam (pair), and the ratio of overhead reduction;
* Samsung [9]: Reference signal overhead (SSB or CSI-RS) and reduced latency for the specific beam management procedure should be considered as performance-KPI for beam prediction in both time and spatial domains.
* Samsung [9]: In this SI, study and evaluate the performance of AI/ML beam prediction in the time domain in terms of the top N beam prediction accuracy and the overhead and latency reduction.
* Beijing Jiaotong University [11]: Overhead, latency and beam selection accuracy of existing L1-RSRP measurement scheme, need to be considered as the benchmark of AI-based beam selection.
* Lenovo [18]: Beam acquisition/selection latency including Tx/Rx beam training should be evaluated for a proposed AI/ML
* Nokia/NSB [19]: RAN1 shall consider the following KPIs for the ML-based beam management use case,
  + Beam management measurement overhead
* NVIDIA [21]: Define a KPI which measures the reference signal overhead for beam management with and without AI/ML based algorithms.
* NVIDIA [21]: Define a KPI which measures the latency for beam management with and without AI/ML based algorithms.
* Qualcomm [23]: The KPIs for temporal beam prediction can be trade-off between reference signal overhead reduction (and/or UE power consumption) and a measure for beam prediction accuracy (such as RSRP prediction accuracy and/or top-M beam selection accuracy) and/or overall system performance (e.g., throughput).
* Qualcomm [23]: The KPIs for spatial (+time) domain beam prediction can be trade-off between reference signal overhead reduction (and/or UE power consumption) and a measure for beam prediction accuracy (such as L1-RSRP prediction accuracy and/or top-M beam selection accuracy) and/or overall system performance (e.g., throughput).

Based on the above proposals, the following question can be discussed.

#### FL1 Medium Priority Question 2-3

**Question 2-3:**

1. Whether the reference signaling overhead reduction ratio can be considered as one of the KPIs for AI/ML in BM (when applicable)? If yes, how to define the metric?
2. Whether latency can be considered as one of the KPI for AI/ML in BM (when applicable)? If yes, how to define the metric?

|  |  |  |
| --- | --- | --- |
| Company | Y/N | Comments |
| Apple | Y | We think both overhead and latency can be considered as optional KPIs, which depend on the use cases. For overhead, metric can be number of required resources for BM, and for latency, the metric can be beam update delay. |
| Nokia, NSB | Y | 1. signaling overhead reduction ratio can be computed as   1-N/M, where N is the number of beam measurements for a subset of beams with size N and M is the number of all beams. When UE specific measurements are needed, N should be specific for each UE, and the metric should account for the number of UEs per sector.   1. Latency may be specific to the packet delay defined as the one-way PDCP layer delay between gNB and UE. |
| Xiaomi | Y | At least the reference signal overhead reduction ratio can be considered as one of the KPIs. And the metric can be the number of RS resource for beam management. |
| vivo | Y | Overhead from beam sweeping and UCI report should be considered as basic KPI. The metric of beam sweeping overhead can be the number of resources for BM in both non-AI algorithm and AI algorithm. The overhead of UCI report can also be reported by companies.  The necessity of considering latency is unclear since we already have RSRP difference as a metric. The necessity and details on how to define latency should be clarified. |
| Intel | Y | Both overhead and latency reduction are important metrics to consider along with beam prediction accuracy and are more important that system performance metrics in the previous section. |
| NVIDIA | Y | 1. Signaling overhead can be measured as the ratio of required number of beams for a given scheme over the number of all beams. With this, signaling overhead reduction can be computed accordingly. 2. AI/ML model inference latency, i.e., the time it takes for the model to make its prediction once it is fed its input. |
| OPPO | Y | RS overhead reduction and latency can be optional KPIs. |
| AT&T | Y | Support overhead reduction and latency metrics. The proposals from Nokia and vivo are good starting points |
| CATT | Y | a) Overhead reduction ratio can be considered as a KPI.  b) Latency reduction seems not applicable to all sub use cases. |
| LGE | Y | Prefer to have overhead reduction ratio as basic KPI. Generalized equation for the ratio can be decided in this meeting. |
| Ericsson | Y | 1. Yes, but the overhead needs to be compared to a state-of-the-art non-ml baseline approach. 2. Unclear what is meant by latency. The latency could comprise of the time to find the strongest beam, however, such information should be visible in the RSRP performance statistics. |
| ZTE, Sanechips | Y | a) At least for the spatial-domain beam prediction, the reference signaling overhead reduction ratio can be considered as one of the KPIs for AI/ML in BM. Otherwise, it may be hard to align performance from different companies since different number of measured partial beams are used as model input. Besides, the number of employed RS resources and corresponding partial beams would also significantly impact the reporting overhead if AI model is deployed at gNB. Intuitively, there should be a trade-off between RS overhead and inference accuracy.  b) Latency shall be clearly defined. For example, the inference latency is strongly dependent on the computing power and software/hardware implementation of gNB/UE, which may be hard to be defined or quantified. |
| CAICT | Y | The detail KPI for latency and overhead reduction could be fixed after baseline performance of non-AI algorithm(s) with comparable accuracy. |
| Samsung | Y for a) | RS overhead is an important KPI. Each company can provide some analysis on the RS overhead. alternatively, we can agree on some reduced RS number for spatial domain prediction, and compare the performance.  The latency here is not very clear. whether this is the reduced latency of measurement/report, or this is latency of AI model inference. Some clarifications are needed. |
| Fujitsu | Y | At least the overhead reduction ratio should be an optional KPI. For the overhead reduction, metric can be the number of measurement resources on DL and reports on UL.  For the latency, it involves more implementation issues and is not recommended as KPI |
| CMCC |  | a) Yes. Nokia's definition is fine.  b) It can be discussed in AI 9.2.1. |
| MediaTek | Y | * Signaling overhead can be considered. * Latency should be defined as the interval between beam indication and UE using the new beam (TCI state). Such latency is usually quite short for intra-cell case (~5ms) but may be much longer for inter-cell beam management due to the need of synchronization. We suggest that latency be considered for inter-cell beam management. |
| HW/HiSi |  | 1. Yes. 2. Open to discuss |
| InterDigital | Y | a) Yes, but RS overhead reduction should be compared with implementation based AI/ML prediction without specification enhancement.  b) Yes, but latency reduction should be compared with implementation based AI/ML prediction without specification enhancement. |
| Lenovo | Y | 1. Yes. Signaling overhead incurred by the AI/ML method, which may include the following, need to be considered as one of the KPIs:    1. Number of reference signals (e.g., CSI-RS, SRS, SS blocks etc.) configured for a beam management procedure.    2. Number of measurement reports and the corresponding report content.    3. Any other signals that need to be exchanged between UE and gNB to support the AI/ML model, such as signaling in another carrier (e.g., FR1), UE location information, spatial/visual features of the environment etc.   Further, the signaling overhead incurred during every phase in the lifecycle of the AI/ML model (training-in case of online training, inference, model update, as applicable) should be reported.  The number of time-frequency resources needed for the signaling required for beam management could be a measure of the signaling overhead.   1. Yes. Beam acquisition/selection latency should be considered as a KPI for a proposed AI/ML model. It should include the following.    1. Tx beam training: The average time required for the UE to acquire/select a gNB Tx beam with a measured L1-RSRP/L1-SINR larger than a threshold among a set of configured Tx beams.   Rx beam training: The average time required for the UE to acquire/select a Rx beam with measurements on the same gNB Tx beam and measured L1-RSRP/L1-SINR larger than a threshold among a set of Rx beams. |
| Qualcomm |  | Yes, to both a) and b). As a representative use case for which these metrics are useful, we can consider wide to narrow beam prediction. Through measurements of wide beams (e.g., SSBs over multiple measurement occasions), we can reduce the search space over narrower beams (e.g., CSI-RS) and therefore reduce the CSI-RS *overhead* as well as beam refinement *latency,* due to sweeping over a smaller number of beams. As we see, for this use case, a metric for overhead as well as latency can appropriately capture the benefits of AI/ML. We can discuss the prospects of exact metric definition in later phases of the study. |

Power consumption were proposed by three companies as the metric for evaluation.

* Huawei/HiSi [1]: Consider the evaluation of the AI/ML based spatial domain beam prediction with the target to obtain accurate narrow beams for CSI-RS beam sweeping to reduce the UE’s overhead and power consumption.
* Ericsson [4]: Use case performance metrics could include metrics defined in Rel-17 EVM for BM including a potential energy saving metric, for example using the models defined in [4].
* CATT [6]: Intermediate KPI: Probability of identifying the best beam (pair), average power loss between the selected beam (pair) and the optimal beam (pair), and the ratio of overhead reduction;
* Qualcomm [23]: The KPIs for temporal beam prediction can be trade-off between reference signal overhead reduction (and/or UE power consumption) and a measure for beam prediction accuracy (such as RSRP prediction accuracy and/or top-M beam selection accuracy) and/or overall system performance (e.g., throughput).
* Qualcomm [23]: The KPIs for spatial (+time) domain beam prediction can be trade-off between reference signal overhead reduction (and/or UE power consumption) and a measure for beam prediction accuracy (such as L1-RSRP prediction accuracy and/or top-M beam selection accuracy) and/or overall system performance (e.g., throughput).
* Ericsson [4]: Evaluation results could include ML-performance metrics related to beam predictions, both comprising classification and regression performance metrics

Based on the above proposals, the following question can be discussed.

#### FL1 Medium Priority Question 2-4

**Question 2-4:**

1. Whether power consumption can be considered as one of the KPIs for AI/ML in BM? If the answer is yes, how to define the power consumption metric?

|  |  |  |
| --- | --- | --- |
| Company | Y/N | Comments |
| Apple | Y | UE power consumption can be considered as an optional KPI, which depend on the use case. |
| Nokia, NSB | Y | Yes, discussion should continue to define more details about the UE power consumption model. |
| vivo | Y | Support. Power consumption can be considered by defining a reference unit, e.g., define single unit power consumption as single resource reception with one RX beam. |
| Intel | N | This can be a secondary consideration once sub-use-cases with definitive performance gain are finalized. |
| NVIDIA | Y | More discussion is required to define a proper power consumption metric. |
| OPPO |  | Not sure how to evaluate the power consumption for AI-based solution. In general, the implementation of AI is device-specific. |
| CATT | N | Actually, power consumption can be reflected by RS measurement (overhead) reduction. It does not have to be a standalone metric. |
| LGE | N | Similar view as Intel. |
| Ericsson | Y | Using the models in Ericsson [4] |
| ZTE, Sanechips |  | We share similar view with Intel and OPPO that more details are needed for evaluating power consumption. |
| CAICT |  | Power consumption is hard to measure. |
| Samsung |  | Need to discuss whether a proper UE power consumption model can be easily found and agreed.  We are open to have some qualitative analysis or observation in general. |
| Fujitsu | N | It’s difficult to get power consumption results through simulation only. |
| CMCC |  | The method to evaluate the power consumption needs to be specified. It can be discussed in AI 9.2.1. |
| MediaTek | N | Can be an optional metric. |
| HW/HiSi |  | We are open to discuss. The issue will be how to model the power consumption and how to align on a common understanding. |
| InterDigital |  | We are fine with the proposal if it is optional. |
| Lenovo | Y | Yes. Power consumption is an important consideration, especially, for UE and it is tightly coupled to computational complexity. It can be measured as the average power consumed for beam alignment/finding procedure. |
| Qualcomm | Y | The metric can be defined as a function of number of measurements that the UE needs to perform for BM. For the use case in which UE only measures a down-sampled set of SSBs to predict strongest SSB ID, the benefit cannot be captured properly through overhead reduction as the overhead remains fixed due to legacy SSB. The benefit can be quantified properly through less UE measurements and hence UE power saving. This illustrative use case exemplifies why we need to consider the UE power saving aspect *in addition to* RS overhead reduction, as there is no overhead reduction in this example. We can discuss the prospects of exact metric definition in later phases of the study. |

## Capability-related KPIs

### 2.2.1 Generalization

Generalization is one of the important aspects to verify the performance of AI/ML model. The following proposals were proposed:

* Huawei/HiSi: [1] For AI/ML-based beam management evaluation, study how to evaluate the generalization of the AI/ML model in simulation.
* Vivo [7]: Generalization performance should be considered as performance-related KPIs.
* Vivo [7]: Different number of Tx beam and Rx beam should be considered for construction of the data set for generalization performance evaluation.
* Samsung [9]: For the use case of AI/ML based beam management, at least the following capability-related KPI shall be considered:
  + Generalization of AI/ML model by considering the performance degradation for test case with mixed scenarios.
* Lenovo [18]: Inference accuracy of a proposed beam management AI/ML model (without model updates) should be reported for different channel models including UMa, UMi (with various parameters), UE speed, UE orientations, and beam codebook design.
* Nokia/NSB [19]: For training/testing the supervised learning ML model, the number of scenarios/deployments and configurations of other simulation parameters can be extended and generalized beyond those detailed in Table 2.1 2.

#### FL1 Medium Priority Question 2-5

**Question 2-5:**

1. Whether generalization should be one of the KPIs for AI/ML in BM?
2. If the answer is yes, how to define or test the generalization performance? For example, how to mix or extend the simulation parameters?

|  |  |  |
| --- | --- | --- |
| Company | Y/N | Comments |
| Apple |  | We are open, but we have the same question as b) |
| Nokia, NSB | Y | Based on table 1 and 2, the ML model trained for one specific scenario can be valid for single sector or multiple sectors. Discussion should continue to define more details and establish a procedure to test the ML mode generalization capabilities. |
| Xiaomi |  | Open to discuss. |
| vivo | Y | Support generalization as a basic KPI in BM.  It should not be forced to realize the same Rx beam number for all accessed UEs, thus model design should consider the output target of an AI model used for beam prediction probably used for different number of Tx beams and Rx beams. Similarly, different drops, scenarios and UE trajectory can also be considered in generalization.  Generalization performance should be one of the basic KPIs for AI/ML in BM.  To define and test generalization performance, it can be easily verified as following:   * Step1: Using a first set of parameters (including e.g., Tx beams, Rx beams, drops, scenarios and UE trajectories) to generate training data set A; * Step2: Using a second set of parameters different from those in step 1 (including e.g., Tx beams, Rx beams, drops, scenarios and UE trajectories) to generate testing set B; * Step3: Testing the performance of the model trained in step1 with the dataset generated in step2. |
| Intel |  | More discussion is needed for how to define generalization. Is training across different channel models and inferencing in one of the models used for training an example of generalization? |
| NVIDIA | Y | Similar to the discussion in 9.2.2.1, a list of configuration parameters and a list of scenarios can be formulated for companies to discuss further. |
| OPPO |  | We are open to discuss this issue. It seems difficult to define an KPI in terms of generalization. But the generalization performance can be evaluated e.g., by using data sets of mixed scenarios. |
| AT&T | Y | This is important, but will require additional scenario definitions (venue, indoor hotspot, etc.) |
| CATT | Y | For generalization evaluation, we can simulate the AI model performance where the training data and testing data are from different scenarios, different gNB/UE antenna configurations, different UE distributions… |
| LGE |  | Open to discuss, but it could be later after the assumption on AI/ML model for BM is stabilized. |
| Ericsson | Y | 1. Yes, this is very important 2. By defining a mix of simulation parameters. For example, training on UEs moving at 60 km/h, and evaluate on UEs moving on 120 km/h. In another example, a model trained from one drop is tested on another UE drop. |
| ZTE, Sanechips |  | Open to discuss. |
| CAICT |  | The generalization of AI/ML model for BM is hard to measure. |
| Samsung |  | Need more discussion on how the generalization is required, e.g, if a different gNB beambook is used, seems it is hard to use existing neural network to be applied in the new case, and anyway the new training is required. Here are several cases, in which generalization should be needed, based on our understanding:   * Different scenario, in which channel modeling is changed. * Different UE trajectory and mobility model. |
| Fujitsu |  | We are open. But at least in initial stage, generalization is not necessary as one of the KPIs. |
| CMCC | Y | We support to discuss it. We believe vivo’s suggestion is a good starting point. The performance of a generalized model (can be used in different parameter settings) should also be evaluated. |
| MediaTek |  | We believe that the trained model should be “generally applicable”, but we would like to have some discussion on what “generalization” means. |
| HW/HiSi |  | Generalization is important and should be studied. But at this stage we think it will be difficult to quantify the KPI. |
| InterDigital |  | We believe that further discussion is needed. |
| Lenovo | Y | We should consider generalization as one of the KPIs to find out whether the model adapts to dynamically changing network environment (and corresponding changes in the statistical characteristics of the data). Some of the different network conditions are indicated below (please note that this is not an exhaustive list).   * 1. Different channel conditions      + Channels with single dominant path vs. multipath   2. Different UE speeds and rotation patterns      + Different possible orientations of the UE, UEs with different speeds etc.   3. Does the AI/ML model work with any generic beam design?   It can be tested by computing all the performance KPIs (e.g., beam prediction accuracy, latency, overhead etc. that are discussed before) of a proposed beam management AI/ML model under different channel models (with various parameters), UE speeds, and different beam designs etc. |
| Qualcomm |  | a) Yes. There are two different ways to address generalization. One approach is to develop a single (potentially more complex) AI/ML model that works across a more diverse set of scenarios *or* develop several (smaller, less complex) AI/ML models and perform model switching for different scenarios.  b) Assuring that the training data is rich and diverse enough is an important aspect, which needs to be ensured for proper generalization performance. There are multiple ways/levels to look into generalization performance. First example is to train on a first set of UEs and test on a second set of UEs without changing simulation parameters across UEs. Another option is to change parameters of a deployment across training and test dataset (within a cell) and evaluate the performance (such as different trajectories and speeds for UE movement for train and test dataset). A third example is to train in a first cell and test on a second cell. As mentioned before, the premise for good generalization is good training data coverage. If we train based on a first set of simulation assumptions and test based on a totally different set of assumptions and the model does not perform well, it is hard to argue against the generalization capability, as the training and test data have totally different distributions. |

### 2.2.2 Model size and computational complexity

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

* Vivo [7]: Beam sweeping overhead, AI processing latency, computational complexity, model size and power consumption can be considered are all relevant for assessment.
* Samsung [9]: For the use case of AI/ML based beam management, at least the following capability-realted KPI shall be considered:
  + Size of AI/ML model;
  + Complexity of training and inference of AI/ML operation.
* Futurewei [12]: Include model/space complexity as one of the KPIs to evaluate the complexity of AI/ML-based approach for beam management use case. Model/space complexity measurements include number of parameters in the model and memory usage.
* Futurewei [12]: Include computational complexity as one of the KPI list to evaluate the complexity of AI/ML-based approach for beam management use case. Computational complexity can be estimated using number of floating-point operations (FLOPs) or number of multiplies and accumulates (MACs).
* CAICT [14]: AI/ML model size should be controlled during the evaluation phase.
* Lenovo [18]: Computation complexity of the proposed AI/ML model should be evaluated at every phase in the model lifecycle-training, inference, update.
* NVDIA [21]: Provide information on the used computing platform (such as the GPU model) when reporting KPIs for AI/ML based beam management.
* Fujitsu [24] The following KPIs are recommended for complexity of AI/ML models.
  + Computational complexity of AI/ML model inference
  + Memory size for AI/ML model storage

#### FL1 Low Priority Question 2-6

**Question 2-6:**

1. Whether AI/ML model size can be considered as one of the KPIs for AI/ML in BM? If the answer is yes, how to quantify it, e.g., reported by each company with model size in Bytes?
2. Whether computation complexity at least for inference can be considered as one of the KPIs for AI/ML in BM? If the answer is yes, how to quantify it, e.g., reported by each company with number of floating-point operations (FLOPs) or number of multiplies and accumulates (MACs)?

|  |  |  |
| --- | --- | --- |
| Company | Y/N | Comments |
| Apple | N |  |
| Nokia | Y | a) We may indicate the model size in Byte as well as the total number of trainable parameters.  b) Yes, we have some details in 9.2.1, FLOPs could be one way to proceed. |
| Xiaomi |  | Open to discuss. |
| vivo |  | Open to discuss |
| Intel | N | At this early phase, we should evaluate absolute gain available from AI/ML models. These should be considered at a later phase. |
| NVIDIA | Y | a) Number of trainable model parameters, model size in Byte  b) Yes, number of floating-point operations (FLOPs) |
| OPPO |  | Companies are encouraged to report the mode size and FLOPs. How to use them can be further studied. |
| CATT |  | Model size and computation complexity should be reported/evaluated anyway, thus natural to be KPI. Details can be further discussed. |
| LGE |  | Open to discuss. |
| Ericsson | Y | 1. Yes, for example number of parameters. We consider model complexity estimations are mainly relevant valid for UE sided models. 2. FLOPs and model execution frequency could be presented. |
| ZTE, Sanechips |  | AI/ML model size as well as the number of FLOPs/MACs can be intermediate metrics for complexity evaluation. Nevertheless, considering the impact of the software environment and hardware implementation, more discuss is needed for quantify the complexity when comparing the strengths and weaknesses of different AI models. |
| CAICT | N | AI/ML model size should be limited for performance comparison. The size and structure of AI/ML model should be reported. |
| Samsung |  | Company can report the model size in Bytes and FLOPs |
| Fujitsu | Y | Model size is reported by companies in Bytes. And the computation complexity is reported as FLOPs |
| CMCC |  | Open to discuss |
| MediaTek | N | AI/ML model size is important from implementation perspective. However, at this early stage we should focus on the gain of introducing AI/ML assistance for BM. |
| InterDigital |  | We are fine with it if it is optional. |
| Lenovo | Y | 1. It is important to take “model size” into consideration. However, we think it can be considered as “memory required for the AI/ML model” as a part of the “Complexity” KPI discussed below. 2. Yes. Computational complexity should be considered as a KPI. We suggest that the complexity KPI should consider detailed measures that capture the hardware & software complexity (including the memory footprint) as shown in the table below (reproduced here from [R1-2204416](https://www.3gpp.org/ftp/TSG_RAN/WG1_RL1/TSGR1_109-e/Docs/R1-2204416.zip)).  |  |  | | --- | --- | |  | **Parameters** | | **Hardware** | Peak floating-point operations per second (FLOPS), | | Peak bandwidth to access the memory (byte/sec) | | Memory for the AI/ML model and input/output data (byte) | | Energy consumption per operation (J/op) | | Penalty of interaction with bus | | Need for measurements from any auxiliary (non-3GPP) hardware components (e.g., sensors/actuators/GPS etc.)? | | **Software** | Number of multiply-accumulates (MACs) | | Number of weights of the neural network | | Number of memory accesses | | Number of bytes per memory access (byte) | | Interaction Operational Intensity (FLOPS/byte) | |
| Qualcomm | Y | a) Yes, model size can be a KPI and can be reported in terms of number of parameters, instead of bytes  b) Yes, in principle. It is important to note that both a) and b) are for information purposes and do not directly translate into the relevant complexity such as power consumption, latency, etc. The real complexity impact is determined by implementation and is merely loosely correlated with the reported KPIs here. |
|  |  |  |

### 2.2.3 Other KPIs

Several other KPs were proposed, including reporting/model management overhead, AI processing latency, scalability:

* Vivo [7]: Beam sweeping overhead, AI processing latency, computational complexity, model size and power consumption can be considered are all relevant for assessment.
* LGE [13]: Study whether the procedure for training/update/management of AI/ML model can be considered as overhead and latency of beam management or not, and how to quantify the overhead/latency of AI/ML model.
* Lenovo [18]: Beam management signaling (e.g., RS, measurement reports) overhead required for a proposed AI/ML model (at every phase in the model lifecycle) should be evaluated. This can be reported relative to state-of-the-art Rel-17 based baseline schemes.
* Lenovo [18]: Inference accuracy sensitivity of a proposed beam management AI/ML model should be evaluated for a range of noisy input data and delay constraints.
* Lenovo [18]: Scalability of a proposed beam management AI/ML model for different parameter settings, e.g., number of beams at gNB/UE, should be evaluated.
* Lenovo [18]: KPIs for the AI/ML beam management evaluation should at least include accuracy and effectiveness, beam measurement overhead, latency, training convergence time, model robustness, scalability, adaptability and complexity/power consumption.
* NVIDIA [21]: Define a KPI which measures the measurement reporting overhead for beam management with and without AI/ML based algorithms.
* Lenovo [18]: A proposed AI/ML model that is based on online learning (e.g., online training, model update/re-training) should evaluate the convergence time for the model to reach a valid inference state.

#### FL1 Low Priority Question 2-7

**Question 2-7:**

1. Any other KPI/metric needs to be considered for AI/ML in BM?

|  |  |
| --- | --- |
| Company | Comments |
| Lenovo | The following relevant metrics may also be considered as KPIs for the use case of BM (as presented in [R1-2204416](https://www.3gpp.org/ftp/TSG_RAN/WG1_RL1/TSGR1_109-e/Docs/R1-2204416.zip)).   1. Side information from e.g., sensor(s) and hardware support   Does the AI/ML method require side information (e.g., measurements from non-3GPP technologies). For example, does the proposed beam management AI/ML method require measurements from additional sensors like GPS, LIDAR, Camera, or additional hardware like memory storage, GPU execution?   * 1. UE   2. gNB   For example, an AI/ML method may require the gNB to have powerful cameras (or equipping a vehicular UE with a LIDAR) so that the BM can be done based on the signals from these sensors. Such an AI/ML method can result in an additional cost and support for sensors, which need to be taken into account while comparing it with other methods.   1. Robustness: Sensitivity of the beam management AI/ML model to 2. Errors in the data (e.g., erroneous measurements exchanged between UE and gNB) 3. Latency (e.g., latency in generating and reporting the measurement reports)   Importance of this KPI may be described by considering the following scenario: Two methods might perform equally well under favorable conditions (such as, highly reliable and low latency reporting of L1-RSRP/L1-SINR measurements over the network). However, performance of one of the methods may degrade substantially with errors in the RSRP/SINR measurements or with an increased latency in the availability of measurement reports at the inference node. It is recommended to evaluate robustness by considering a range of signaling error probabilities and latency values in the low, moderate and high SNR/SINR regimes.  3. Scalability  This KPI reveals whether the model scales well if the network parameters change. AN indicative list of the relevant parameters to be considered here for beam management use case are  a. Number of beams at gNB and UE  b. Number of active UEs in a multi-beam scenario (when the gNB can simultaneously form more than one beam)  c. UE mobility  We think scalability needs to be included as a KPI due to the following reasons: The inference complexity of a method may increase linearly or exponentially with the cardinality of the set of the beams to be searched over. A method may maintain the same complexity but its accuracy may degrade significantly with the increase in the search space. Some methods might work well under stationary/slowly changing environments and may significantly underperform in a highly dynamic environment. |
| Qualcomm | Other KPIs can be considered such as overhead needed for model management (frequent model switching due to poor generalization), overhead needed for data collection at gNB, etc. Some of these KPIs may be difficult to quantify, but the possibility of considering other KPIs should not be precluded and can be up to companies. |

## Baseline performance

Some companies provided some analysis on baseline performance for benchmark.

* CATT [6]: On the simulations for AI/ML-based beam management, conventional beam management strategy (i.e. exhaustive beam sweeping) can be considered as baseline.
* InterDigital [5]: 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.
* Samsung [9]: EVM on AI/ML based beam measurement feedback compression shall at least be able to evaluate the system performance while considering one or both of the following aspects
  + (1) the compressed beam measurement feedback has lower feedback overhead as compared to the legacy feedback for a given number of reported beams
  + (2) the number of reported beams in the compressed beam measurement feedback is larger than the legacy feedback for the same beam measurement feedback overhead.
* OPPO [10]: To make more meaningful comparison between traditional beam selection scheme and AI/ML beam prediction, study and decide the EVM and benchmark for comparison.
* CMCC [16]: Baseline of spatial domain beam prediction needs to be studied, the following three options of baseline can be considered:
  + Option 1: gNB performs exhaust beam sweeping, UE selects best beam pair among all beam pairs.
  + Option 2: gNB performs sparse beam sweeping with fixed sparse pattern, UE selects best beam pair among measured beam pairs.
  + Option 3: gNB performs sparse beam sweeping with variable sparse pattern, UE selects best beam among measured beam pairs.
* Nokia/NSB [19]: Baselines for supervised learning spatial-temporal domain beam prediction are needed. The following are possible candidates:
  + UE measures all the CSI-RS resources with an exhaustive search at all time instants and selects the best beam at each time instant.
  + UE selects the best beam based on the measurements of all the CSI-RS resources during the time instants within the observation window and maintains the selected beam during the time instants of the prediction window.
* Intel [20]: Baseline performance evaluation schemes for benchmarking the performance of AI/ML tools should be based on exhaustive beam search and more practical hierarchical beam search procedures.
* Qualcomm [23]: Based on the agreed KPIs, RAN1 should study and identify baseline prediction schemes for temporal beam prediction as a benchmark for AI/ML-based schemes to compare against
* Qualcomm [23]: Based on the agreed KPIs, RAN1 should study and identify baseline prediction schemes for spatial domain beam prediction as a benchmark for AI/ML-based schemes to compare against

The baseline performance is subject to each sub-use case. However, this is one of the important issues for evaluation, we can have some open discussion

#### FL1 Medium Priority Question 2-8

**Question 2-8:**

1. For spatial domain beam prediction, what can be the baseline performance?
2. For time domain beam prediction, what can be the baseline performance?

Note: The baseline performance of other sub-use cases can be discussed after the sub-use cases are well defined.

|  |  |
| --- | --- |
| Company | Comments |
| Apple | a) L1-RSRP from ideal beam, beam selection from non-AI scheme (spatial correlation), and random beam  b) L1-RSRP from ideal beam, and the latest beam (assuming no beam change) |
| Nokia, NSB | * + For spatial domain beam prediction:   **Alternative 1a :**  UE measures all CSI-RS resources  **Alternative 2a :**  UE measures a subset of CSI-RS resources with given size  **Alternative 3a :**  UE measures all SSB resources   * + For time domain beam prediction   **Alternative 1b :** UE measures all the CSI-RS resources with an exhaustive search at all time instants and selects the best beam at each time instant.  **Alternative 2b :** UE selects the best beam based on the measurements of all the CSI-RS resources during the time instants within the observation window and maintains the selected beam during the time instants of the prediction window. |
| Xiaomi | For spatial domain, beam selection from measurement on sparse beams/wide beams can be the baseline. |
| vivo | Those best measured RSRP can be directly as baseline for comparison, as we have done for Rel-15/16/17 design. |
| Intel | The baselines are dependent heavily on sub-use-case definitions. Suggest considering this after sub-use-cases are finalized. |
| NVIDIA | 1. Upper bound: Genie (best beam); Lower bound: UE measures a (random/fixed) subset of beams 2. Upper bound: Genie (best beam); Lower bound: latest best beam in the observation window applied to the prediction window |
| OPPO | Since there is no common baseline for beam management, the baseline scheme can be reported by each company in their evaluation. In general, company can use any non-AI scheme as the baseline. |
| CATT | At least for spatial domain prediction, exhaustive beam sweeping scheme can be set as the baseline. |
| LGE | For spatial domain beam prediction, exhaustive beam search can be the baseline.  For time domain beam prediction, specific periodicity of beam reporting can be decided for baseline performance. |
| Ericsson | a) L1-RSRP from ideal beam, and beam selection from non-AI scheme such as selecting K closest beams in the spatial domain (adjacent beams)  b) Using the latest strongest measured beam in the prediction window could be used. |
| ZTE, Sanechips | We share similar view with Nokia. |
| CAICT | a) gNB performs exhaust beam sweeping, UE selects best beam pair among all beam pairs.  b) the latest beam could be used as baseline. |
| Samsung | a) Upper bound: Genie aided approach by assuming all beams being measured. Baseline: Only measure the restricted subset of beambook.  b) Genie assuming knowledge of the optimal beam pair in the predicted time. We are open to discuss the latest beam, but how to quantify top-N beam accuracy is should be discussed since it is unclear for the latest beam. |
| Fujitsu | For spatial domain beam prediction, the exhaustive beam search can be used to compare the gains of overhead reduction of AI/ML model. And the hierarchical beam search can be used to compare the performance gains of AI/ML model. |
| CMCC | a)   * Option 1: best beam pair among all beam pairs. * Option 2: best beam pair among a fixed subset of all beam pairs. * Option 3: best beam pair among a random subset of all beam pairs.   b) At least consider the followings   * Option 1: latest best beam pair with the same overhead with AI-based method. * Option 2: latest best beam pair with a frequent beam measurement and beam reporting. This can be viewed as the upper bound of traditional scheme. |
| MediaTek | The baseline performance depends largely on sub-use-cases. However, L1-RSRP can be considered as a general baseline. |
| HW/HiSi | Since there is not standardized baseline in legacy BM, we think it is up to company to report their baseline. One example for spatial BM could be to use the optimal beam from the SSB sweeping phase and then sweep its neighboring narrow beams. For time domain beam prediction, we are also open to discuss further. |
| InterDigital | For both time domain and spatial domain, optimal beam selection ratio, L1-RSRP difference and system performance should be considered. |
| Lenovo | a) The baseline scheme can be genie-aided ideal beam (equivalently, beam identified from exhaustive search). In addition to the ideal beam, each company be allowed to consider a state-of-the-art non-AI/ML scheme and report it while presenting their performance results.  b) Genie-aided ideal beam (equivalently, beam identified from exhaustive search). Additionally, each company be allowed to consider a state-of-the-art non-AI/ML scheme and report it while presenting their performance results. |
| Qualcomm | In our view, for both a) and b), it is important to identify state-of-the-art *competitive* baselines to compare against. Comparing AI/ML performance to exhaustive search or random beam selection may potentially lead to exaggerated gains (given the KPIs: top-K beam selection accuracy and/or overhead reduction) that may not highlight the actual benefit of AI/ML-based approaches and may lead to misinterpreted results/conclusions. This state-of-the-art baseline can be defined for each use case, and the prospect of other prediction-based baselines (non-AI/ML) could also be discussed. |

# Others

Some companies suggest to consider multiple scenarios for evaluations.

* InterDigital [5]: Consider the following AI/ML aided beam prediction scenarios for evaluations.
  + Simple specification extension of UE reporting which enables AI/ML beam prediction
  + Partial beam measurement
  + Beam prediction in time domain
  + Association between beams with different beam widths
  + Handling of UE Rx beams
* Qualcomm [23]: Based on evaluations using the agreed KPIs, RAN1 should identify the deployments/scenarios/use cases for which AI/ML based temporal beam prediction may be beneficial.

# Appendix: Detailed evaluation assumptions

Table 1 Baseline assumptions for SLS

|  |  |
| --- | --- |
| Parameters | Values |
| Frequency Range | FR2 @ 30 GHz,   * SCS: 120 kHz * BW: 80 MHz |
| Scenarios | High speed @FR2:   * Dense Urban (macro-layer only, TR 38.913) @FR2, 200m ISD, 2-tier model with wrap-around (7 sites, 3 sectors/cells per site), 100% outdoor   One UE is dropped for each of the 21 sectors/cells (see mobility description below) |
| UE Speed | For Dense Urban: 60 km/hr and 120 km/hr |
| 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 | (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  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 | Number/location of panels: 3 panels (left, right, and back)  Panel structure: 1x4x2 or (M, N, P) = (1, 4, 2), dH = 0.5 λ  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 | Full buffer |
| 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 |
| Algorithm details (when applicable) | Companies to report:   * Beam reporting mechanism * Beam metric L1-RSRP * Number of active panels |
| Other potential impairments | Not modelled (assumed ideal).  If impairments are included, companies will report the details of the assumed impairments |

Table 2 Baseline assumptions for SLS: Additional simulation assumptions for Dense Urban scenario (FR2) mainly from TR 38.802 Table A.2.1-1, and TR 38.901.

|  |  |
| --- | --- |
| Parameters | Values |
| Carrier Frequency | 30 GHz |
| Scenario | UMa LOS |
| System BW | 80 MHz |
| 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 |

# Reference

|  |  |  |  |
| --- | --- | --- | --- |
| [1] | [R1-2203142](https://www.3gpp.org/ftp/TSG_RAN/WG1_RL1/TSGR1_109-e/Docs/R1-2203142.zip) | Evaluation on AI/ML for beam management | Huawei, HiSilicon |
| [2] | [R1-2203250](https://www.3gpp.org/ftp/TSG_RAN/WG1_RL1/TSGR1_109-e/Docs/R1-2203250.zip) | Evaluation assumptions on AI/ML for beam management | ZTE |
| [3] | [R1-2203255](https://www.3gpp.org/ftp/TSG_RAN/WG1_RL1/TSGR1_109-e/Docs/R1-2203255.zip) | Model and data-driven beam predictions in high-speed railway scenarios | PML |
| [4] | [R1-2203283](https://www.3gpp.org/ftp/TSG_RAN/WG1_RL1/TSGR1_109-e/Docs/R1-2203283.zip) | Evaluations on AI-BM | Ericsson |
| [5] | [R1-2203374](https://www.3gpp.org/ftp/TSG_RAN/WG1_RL1/TSGR1_109-e/Docs/R1-2203374.zip) | Discussion for evaluation on AI/ML for beam management | InterDigital, Inc. |
| [6] | [R1-2203453](https://www.3gpp.org/ftp/TSG_RAN/WG1_RL1/TSGR1_109-e/Docs/R1-2203453.zip) | Discussion on evaluation on AI/ML for beam management | CATT |
| [7] | [R1-2203552](https://www.3gpp.org/ftp/TSG_RAN/WG1_RL1/TSGR1_109-e/Docs/R1-2203552.zip) | Evaluation on AI/ML for beam management | vivo |
| [8] | [R1-2203810](https://www.3gpp.org/ftp/TSG_RAN/WG1_RL1/TSGR1_109-e/Docs/R1-2203810.zip) | Evaluation on AI/ML for beam management | xiaomi |
| [9] | [R1-2203899](https://www.3gpp.org/ftp/TSG_RAN/WG1_RL1/TSGR1_109-e/Docs/R1-2203899.zip) | Evaluation on AI ML for Beam management | Samsung |
| [10] | [R1-2204017](https://www.3gpp.org/ftp/TSG_RAN/WG1_RL1/TSGR1_109-e/Docs/R1-2204017.zip) | Evaluation methodology and preliminary results on AI/ML for beam management | OPPO |
| [11] | [R1-2204059](https://www.3gpp.org/ftp/TSG_RAN/WG1_RL1/TSGR1_109-e/Docs/R1-2204059.zip) | Evaluation methodology of beam management with AI/ML | Beijing Jiaotong University |
| [12] | [R1-2204102](https://www.3gpp.org/ftp/TSG_RAN/WG1_RL1/TSGR1_109-e/Docs/R1-2204102.zip) | Discussion on evaluation of AI/ML for beam management use case | FUTUREWEI |
| [13] | [R1-2204151](https://www.3gpp.org/ftp/TSG_RAN/WG1_RL1/TSGR1_109-e/Docs/R1-2204151.zip) | Evaluation on AI/ML for beam management | LG Electronics |
| [14] | [R1-2204182](https://www.3gpp.org/ftp/TSG_RAN/WG1_RL1/TSGR1_109-e/Docs/R1-2204182.zip) | Some discussions on evaluation on AI-ML for Beam management | CAICT |
| [15] | [R1-2204240](https://www.3gpp.org/ftp/TSG_RAN/WG1_RL1/TSGR1_109-e/Docs/R1-2204240.zip) | Evalulation on AI based Beam Management | Apple |
| [16] | [R1-2204297](https://www.3gpp.org/ftp/TSG_RAN/WG1_RL1/TSGR1_109-e/Docs/R1-2204297.zip) | Discussion on evaluation on AI/ML for beam management | CMCC |
| [17] | [R1-2204377](https://www.3gpp.org/ftp/TSG_RAN/WG1_RL1/TSGR1_109-e/Docs/R1-2204377.zip) | Discussion on evaluation on AI/ML for beam management | NTT DOCOMO, INC. |
| [18] | [R1-2204419](https://www.3gpp.org/ftp/TSG_RAN/WG1_RL1/TSGR1_109-e/Docs/R1-2204419.zip) | Evaluation on AI/ML for beam management | Lenovo |
| [19] | [R1-2204573](https://www.3gpp.org/ftp/TSG_RAN/WG1_RL1/TSGR1_109-e/Docs/R1-2204573.zip) | Evaluation on ML for beam management | Nokia, Nokia Shanghai Bell |
| [20] | [R1-2204795](https://www.3gpp.org/ftp/TSG_RAN/WG1_RL1/TSGR1_109-e/Docs/R1-2204795.zip) | Evaluation for beam management | Intel Corporation |
| [21] | [R1-2204842](https://www.3gpp.org/ftp/TSG_RAN/WG1_RL1/TSGR1_109-e/Docs/R1-2204842.zip) | On evaluation assumptions of AI and ML for beam management | NVIDIA |
| [22] | [R1-2204862](https://www.3gpp.org/ftp/TSG_RAN/WG1_RL1/TSGR1_109-e/Docs/R1-2204862.zip) | Evaluation methodology aspects on AI/ML for beam management | AT&T |
| [23] | [R1-2205026](https://www.3gpp.org/ftp/TSG_RAN/WG1_RL1/TSGR1_109-e/Docs/R1-2205026.zip) | Evaluation on AIML for beam management | Qualcomm Incorporated |
| [24] | [R1-2205078](https://www.3gpp.org/ftp/TSG_RAN/WG1_RL1/TSGR1_109-e/Docs/R1-2205078.zip) | Evaluation on AI/ML for beam management | Fujitsu Limited |
| [25] | R1-2205102 | AI-assisted Target Cell Prediction for Inter-cell Beam Management | MediaTek Inc. |