**3GPP TSG-RAN WG2 #126 *R2-24xxxxx***

**Fukuoka, Japan, May 20th -24th, 2024**

Agenda Item:

Source: OPPO

Title: [POST125bis][021][AIML mobility] Simulation assumptions and methodology (OPPO)

Document for: Discussion, Decision

# Introduction

This document is to address the following email discussion:

* [POST125bis][021][AI/ML mobility ] Simulation assumptions and methodology (Oppo)

Intended outcome: Agree to set of common and RRM prediction use case simulation assumptions and methodology

Deadline: three weeks

|  |  |  |
| --- | --- | --- |
| Company name | Delegate name | Email address |
| OPPO | Zhongda Du | [duzhongda@oppo.com](mailto:duzhongda@oppo.com) |
| NTT DOCOMO | Xin Wang | [wangx@docomolabs-beijing.com.cn](mailto:wangx@docomolabs-beijing.com.cn) |
| Apple | Sasha Sirotkin | ssirotkin@apple.com |
| Mediatek | Yuanyuan Zhang | Yuany.zhang@mediatek.com |
| Samsung | Sangkyu Baek | sangkyu.baek@samsung.com |
| vivo | Xiang Pan | panxiang@vivo.com |
| Ericsson | Cecilia Eklöf | [cecilia.eklof@ericsson.com](mailto:cecilia.eklof@ericsson.com) |
| Xiaomi | Xing Yang | Yangxing1@xiaomi.com |
| CMCC | Fang Xie | xiefang@chinamobile.com |
| ZTE | LiuJing | liu.jing30@zte.com.cn |
| Nokia | Endrit Dosti | endrit.dosti@nokia.com |
| Intel | Ziyi Li | ziyi.li@intel.com |
|  |  |  |

# Discussion

## General aspect

Before diving into discussion on detailed simulation assumption and methodology/metrics, it could be helpful to identify study goal(s) of this SID in general, which could be used to guide the discussion. For discussion purposes, the case where no AI/ML model is used is called the benchmark case, where measurement is performed based on the current procedure and no measurement is reduced in any domain.

From the online discussion, measurement overhead reduction is one of the most interested goals. It was agreed that it is applied for FR1\_to\_FR1 handover scenario including intra-frequency and inter-frequency measurement and prediction. It could be also applied for FR2\_to\_FR2 intra-frequency measurement and prediction. For such purpose, simulation assumptions could be set as such that measurement could be challenging. For example, when setting the 2nd central frequency for FR1\_to\_FR1 scenario, we’d better choose two frequencies with some distance so that in general measurement gap is needed to perform inter-frequency measurement.

Another study goal could be improvement of handover performance. For FR1\_to\_FR1 case it may be not so attractive considering the handover performance is actually good in the field. But for FR2\_to\_FR2 it does. It is mainly because usually the ISD of FR2 is relatively smaller compared to FR1 cell and it demands effective measurement in order to make quick and right handover decision. On the other hand, UE need to spend time to measure analogy beams which is not efficient compared to FR1 carriers. When setting up simulation assumption for such study goal, some of the parameters like high UE speed could be considered so that handover performance gain powered by AI/ML model can be reflected in some way. You can find more in contributions [16][17] about such study goal discussion.

**Conclusion 1: 1st study goal of evaluation is to reduce measurement overhead**

**Conclusion 2: 2nd study goal of evaluation is to enhance handover performance**

**Question 2.1-1: Do you agree that FR2\_to\_FR2 intra-frequency scenario could be evaluated also to reduce measurement overhead i.e.,1st study goal apart from FR1\_to\_FR1 scenario?**

|  |  |  |
| --- | --- | --- |
| Company | Position: yes or no | comments |
| NTT DOCOMO | Yes |  |
| OPPO | Yes |  |
| Apple | Yes |  |
| Mediatek | Yes | The measurement overhead can be defined in a more general way to consider aspects such the RS transmission, measurement gap, and UE measurement effort. |
| Huawei, HiSilicon | Yes, second priority | We think it is worth evaluating measurement prediction in FR2, but we can do it as a second priority, after seeing the initial results for FR1. |
| Samsung | No (lower priority) | We think FR2 is challenging for HO-related mobility performance. It would be better to focus on performance improvement, and work on overhead reduction later. |
| vivo | Yes with comments | In order to reduce the simulation workload, we suggest that FR1\_to\_FR1 can be the baseline scenario for measurement reduction evaluation, and FR2\_to\_FR2 or FR1\_to\_FR2 can be considered as optional scenarios and companies may provide corresponding simulation results. |
| Xiaomi | Yes |  |
| CMCC | Yes | For FR2 scenario, the potential purpose of the study is to improve the HO performance and reduce measurements |
| ZTE | No with comments | To be honest, it is unclear to us why the answer to this question matters?  According to the RAN2 agreements, for intra-frequency prediction in FR2-FR2, it contains temporal-domain prediction and spatial-domain prediction (including intra-cell and inter-cell), many sub use cases are involved, some of them involving measurement reduction, and some of them don’t.  We tend to agree with Samsung that for FR2 intra-frequency, it’s better to focus on handover performance improvement first, measurement reduction can be considered later. |
| Nokia | Yes as second priority | Maybe ok to consider at a more mature phase of the SI. Prefer to not consider as a starting point. |
| Intel | Yes | In our understanding, the key point of 1st study goal is focusing on the gain for intra-frequency (measurement reduction) and inter-frequency scenario (measurement gap reduction).  From measurement overhead reduction point of view, we think study of FR2\_to\_FR2 has similar priority as FR1\_to\_FR1. Hence, companies can provide simulation results for either scenario to prove meauremnt overhead reduction gain in inter-freq and intra-freq. |

**Question 2.1-2: Do you agree that only FR2\_to\_FR2 intra-frequency scenario is evaluated to improve handover performance i.e., 2nd study goal?**

|  |  |  |
| --- | --- | --- |
| Company | Position: yes or no | Comments |
| NTT DOCOMO | Yes |  |
| OPPO | Yes |  |
| Apple | Not necessarily | System level (e.g. HO performance related) KPIs have not been agreed in RAN2#125bis.  So it’s OK to consider it as a “goal”, as long as it doesn’t become a KPI. But then the question becomes what’s the goal of such a “goal”? |
| Mediatek | No | As an initial step, we should consider evaluating the FR2\_to\_FR2 intra-frequency scenario, as it has the potential to yield considerable benefits for enhancing handover performance. However, we anticipate that RRM prediction could bring about improvements in the FR1\_to\_FR1 scenario, even if the gains might be marginal. It would be beneficial to ensure that our TR includes a comprehensive evaluation for both the FR2 and FR1 scenarios. |
| Huawei, HiSilicon | Yes, but see the comments | We think that the main goal of RRM measurement prediction is to reduce the measurement overhead while maintaining the HO performance or analysing the trade-off between measurement reduction and HO performance degradation. The measurement prediction can of course help to also improve HO performance, but that will be done once the measurement predictions will be translated into measurement event predictions, which we agreed to study after making progress on measurement prediction. However, we agree that once we start evaluating measurement event prediction, we should focus on FR2 where HO performance is challenging. There is no need of improving HO performance in FR1. |
| Samsung | Yes | We think FR2 is challenging for HO-related mobility performance, so we think AI could help improve it. |
| vivo | Yes | Comparing with FR1, the mobility robustness in FR2 is much challenging. Thus the simulation of FR2 should focus on HO performance improvement. |
| Ericsson | Yes | Agree with Huawei. |
| Xiaomi | Yes with comments | The evaluation can focus on FR2, which may show noticabe gain. However, we understand the potential solution can be used in both FR1 and FR2. |
| CMCC | Yes | Same view with Q2.1-1 |
| ZTE | Yes | Agree with Huawei. |
| Nokia | Yes |  |
| Intel | Yes after evaluating RRM measurement prediction intermediate KPIs | For RRM measurement prediction, it is straightforward to know how HO performance could become – if RRM measurement prediction accuracy is low, the HO decision made based on the predicted measurement will not have a comparable good performance.  If the prediction accuracy is good enough, the potential gain in HO performance can be evaluated based on measurement event prediction, where predicted measurements are sent to network based on predicted measurement event at UE-side for handover decision. For such case, it would be valuable to evaluate HO performance. Further, considering HO performance in FR1 is good based on operator’s feedback, we can focus on FR2 for evaluation. |

In order to achieve 1st goal, it is natural that the input measurement results, regardless it is L1 beam level measurement or L3 cell level measurement, should be reduced compared to benchmark case. But we also need to keep in mind that even though measurement can be reduced to some extent, still handover performance should not be degraded too much. It is still FFS whether handover performance will be evaluated directly.

In order to achieve 2nd gaol with maximum performance gain, it is also natural that there is no any reduction of measurement result as input. However, technically it is possible that some level of handover performance can be still achieved as long as measurement is reduced not so much. Then one question could be raised: should we also study those middle cases (with question mark) as illustrated in Figure 2.1-1:



Figure 2.1-1 handover performance vs measurement reduction [17]

Rapporteur’s view is that we should focus on the case where maximum handover performance can be achieved and hence no measurement should be reduced. The middle case could still happen in field by deployment, but there is not so much value for study and evaluation, after cases with the highest gain and the least gain are evaluated.

**Question 2.1-3: For the evaluation exercise for 2nd study goal, do you agree that RAN2 should initially focus on the case with the highest gain and hence input measurement results of AI/ML model is not reduced?**

|  |  |  |
| --- | --- | --- |
| Company | Position: yes or no | Comments |
| NTT DOCOMO | Yes | Our opinion is that the existing measurement should be the baseline for both AI/ML and legacy HO. With initial results of baseline schemes, we can further explore the methods for the 1st study goal. |
| OPPO | Yes |  |
| Apple | Neither, see comments | RAN2 have not agreed to evaluate HO performance, but this question is phrased as if the opposite is true. |
| Mediatek | Yes | Rephrased as ‘RAN2 should initially focus on the case with the highest gain in HO performance without presuming a reduction in measurement overhead’? |
| Huawei, HiSilicon | No, but see the comments | We understand that HO performance improvement will be studied as part of measurement event prediction use case and for this case, we agree we can start with “no measurement reduction” case. But we should also evaluate how the reduction in measurement effort impacts the HO performance. |
| Samsung | Yes | Agree with the rapporteur. |
| vivo | Yes | For the 2nd goal, the most important thing for our study should be to explore the highest gain of AI/ML for handover performance improvement. Therefore, measurement reduction should not be considered in this case. |
| Ericsson | Yes | For the 2nd study goal, in this study item we should initially consider the highest gain case, without reducing the input measurements. |
| Xiaomi | Comments | We understand the input is up to model and solution design. It’s unnecessary to limit the model input. The performance may not be reduced if high accuracy can be provided by AI with reduced input measurement results. We can focus on the performance evaluation at current phase. |
| CMCC | Yes with comments | We could start from the case with the highest gain, the middle cases could be considered if time allowed. |
| ZTE | See comments | Maybe it is better to clarify whether “no measurement reduction” means only temporal-domain prediction Case A with sliding observation window?  If it is for temporal-domain prediction Case A, we think there are two phases:  Phase 1: (sliding window): To evaluate how long the **prediction window** could be, this is done based on the assumption that UE keeps measuring all the time, and observation window is sliding.  Phase 2: (non-sliding window): To further evaluate whether measurement reduction can be done by non-sliding observation window.  In our view, Phase1 can be studied first. |
| Nokia | Yes | Agree this should be the starting point. Improving the HO performance would likely benefit both the NW and UE performance. |
| Intel | Yes | For 1st goal, with measurement reduction, the HO performance should not be degraded. For 2nd goal, it is used to evaluate the best handover performance that AI/ML can achieve (perhaps only using temporal domain prediction), e.g. whether it’s better than legacy or not. |

## Methodology and metrics

### Metrics

To facilitate the discussion, rapporteur takes the liberty to categorise RRM measurement prediction as follows based on [agreements in RAN2#125bis](#_Annex2_agreements_in):

RRM sub case 1: To predict beam level results, then generate cell level results based on the predicted beam results

RRM sub case 2: To directly predict cell level results based on cell level results

RRM sub case 3: Directly predict cell level results based on beam level results

The RRM sub case 1 is bit different from other two sub cases i.e., the direct output of model is beam level results but not cell level result. For RRM case 1, the RSRP difference could be interpreted as L1 RSRP difference for RRM case 1. For RRM case 2 and case 3, RSRP difference can be only interpreted as L3 RSRP difference. It is rapporteur’s understanding here the term “cell level results” refer to L3 cell level measurement results but not L1 cell level measurement results. Without aligned metrics, it will be difficult to compare model performance among 3 RRM sub-cases. Rapporteur believe it is necessary to align metrics among 3 RRM sub cases. Considering the eventual output of the 3 sub cases are all L3 cell level measurement result, the RSRP difference should be interpreted as RSRP difference between predicted L3 cell level measurement result and actual L3 cell level measurement result. Actual measurement is performed in benchmark case.

**Question 2.2.1-1: Do you agree that the prediction accuracy metric for RRM measurement prediction is defined as “RSRP difference between predicted L3 cell level measurement result and actual L3 cell level measurement result” for all RRM sub cases? If you have different interpretation, please provide your version.**

|  |  |  |
| --- | --- | --- |
| Company | Position: yes or no | comments |
| NTT DOCOMO  (Updated comments) | No | We prefer to follow the definition from Rel. 18 AI/ML for BM in TR38.843 with revisions from beam to cell, which is  The difference between the L3 cell-level measurement results (measured on the predicted time slots) of the Top-1 predicted cell and the L3 cell-level measurement results of the actual Top-1 cell.  The reason is to highlight the impacts of the prediction error if the HO target is selected based on the cell with the best-predicted measurement results.  Rapporteur: I am bit confused by the wording “top-1”. Do you assume cluster approach here? If we do it based on either intra or inter-cell case, the cell of input and output is fixed, right?  DCM: Thank you for your comments. In our opinion, the measurement prediction should involve multiple cells to enhance the mobilty performance. The predicton can be either jointly with the cluster approach or separately with the per-cell approach. After that, these predictions may be used for the purpose such as target cell selction. Therefore, it is more important that the RSRP prediction can accurately predict the values of those strong cells, which is quite similar to the Rel. 18 beam management study. The final HO performance will also be more impacted by the prediction accuracy of strong cells.  During Rel. 18, how to define the RSRP gap for beam mangememt study was discussed for a long time and finaly agree the definitions as *“The difference between the ideal L1-RSRP of Top-1 predicted beam and the ideal L1-RSRP of the Top-1 genie-aided beam”,* and other KPIs are not precluded and can be reported.  Considering different companies may have different approaches, we suggest to keep both definitions as the following proposal,   * L3-RSRP difference for measurement prediction   + RSRP difference between predicted L3 cell level measurement result and actual L3 cell level measurement result * L3-RSRP difference for Top-1 predicted cell   + RSRP difference between the L3 cell-level measurement results (measured on the predicted time slots) of the Top-1 predicted cell and the L3 cell-level measurement results of the genie-aided Top-1 cell. * Companies can report one [the first one] or both of them according to their schemes.   [Mediatek] Docomo appears to be examining a scenario where, in addition to the RSRP value, the Top-1 cell is predicted. Consequently, the RSRP difference is characterized as the discrepancy in Layer 3 cell-level results between the predicted Top-1 cell and the actual Top-1 cell.  However, I guess that this KPI may be applicable to other use cases, such as event prediction. In the context of RRM prediction, we have yet to explore subsequent steps on how to leverage these results for target cell selection. Therefore, it might be proper to consider the RSRP difference between the predicted value and the actual value for the same cell for the use case of RRM prediction.  DCM: Thank you for comments. We agree that our suggestion considers a further step after the measurement prediction. As discussed above, the measurement predictions will finally be used to enhance the mobility performance, such as measurement event generation or target cell selection. Therefore, the relative accuracy for those strong RSRP cells is more important than the overall RSRP gap for mobility performance. After a UE connects to the target cell selected based on the prediction, its link quality and system performance are more impacted by our suggested KPI. Therefore, we can have prelimanary observations on how much the predicted measurement can help the mobility before we start the time-consuming system-level HO performance evaluations.  Please also check if our reply to the Rapporteur. Thank you!  For Case 1, the RSRP difference can also be calculated at beam-level.  Rapporteur: this issue is addressed in question 2.2.1-2  DCM: Thank you for your remining. |
| OPPO | Yes | First of all, aligned metric is critical to compare the performance among 3 sub cases. Secondly the eventual output of the algorithm is L3 cell level measurement result, it is very natural to use L3 cell level RSRP. |
| Apple | Yes  (also see comments) | Agree with the definition as proposed by the rapporteur. Also agree that a single metric is needed, as “to consider” these options doesn’t mean we end up using all of them – some downselection is needed. |
| Mediatek | Yes for case 2/3 | It is the RSRP difference between predicted L3 cell level measurement result and actual L3 cell level measurement result (of the same cell). |
| Huawei, HiSilicon | Yes | We agree we need to be able to compare the results. To do so we therefore need to agree on setting of beam consolidation/selection parameters which are used to derive cell level RSRP from L1 beam measurements.  However, we have a general comment on the RRM sub cases. This focus only on cell level measurements and L3 beam level measurements are missing here and in the whole e-mail discussion. L3 beam level predictions are also useful for HO (e.g. for beam selection for initial access). Current agreements clarify that we should study both cell level and L3 beam level predictions, but this is not considered in the discussion thus far:   1. We will consider intra-frequency intra and inter-cell spatial domain measurement predictions, for beam and cell level measurements.   For temporal domain measurement prediction, we will consider the AI-PHY beam management Case A and Case B from the RAN1 AI/ML PHY TR and it applies to both beam level and cell level. As baseline we will focus on pure temporal predicition. |
| Samsung | Yes | The SID explicitly indicate that the scope of this study is “for network triggered L3-based handover.” This means that L3 measurement should be the baseline, at least for RRM prediction. |
| vivo | Yes with comments | The sub-cases are for cell-level measurement prediction. Beam-level prediction is also in the scope. Thus the definition should be restricted to cell-level prediction. Besides, the RRM prediction of neighbour cell is in the scope, The wording can be refined as “For cell-level prediction, the accuracy metrics include RSRP difference between predicted L3 cell level measurement result and actual L3 cell level measurement result for the same cell” |
| Ericsson | Yes | For sub-case 1 we need to agree how to derive cell quality from beam measurements and predictions. |
| Xiaomi | Yes | According to the agreement, AI output only provides cell level measurement result. Cell ranking, i.e. top K, is not the direct output of AI model. Therefore, RSRP difference is enough. |
| CMCC | Yes (see comments) | We think L1-RSRP difference is also necessary for RRM sub case 1. |
| ZTE | Yes with comments | We have different view on what Rapp said that case 1 uses “L1 cell level measurement results”, because case 1 also follows the legacy cell quality derivation procedure. In our view, all the three use cases referring to L3 cell level measurement results.  Instead of discussing L1 or L3 cell level measurement results, we think it is worth clarify whether L3 filtering is considered in the simulation. And we think it’s better not to consider L3 filtering in (at least) RRM prediction simulation, so that we can have more accurate assessment on the RSRP prediction accuracy.  More detailed, we suggest to not consider L3 filtering, or consider ki = 0 in L3 filtering.  ***F*n = (1 – *a*)\**F*n-1 + *a*\**M*n**  ***a*** = 1/2(***ki***/4), where ***ki*** is the *filterCoefficient* |
| Nokia | Yes | For cases 2, 3 this can be the starting point. For Case 1, similar view as Ericsson. |
| Intel | Yes | In general, we agree to consider RSRP difference between predicted L3 cell level measurement result and actual L3 cell level measurement result of the same cell .  For sub case 1, we need to align all L3 filtering related parameters to see calculate L3 cell level measurement based on measured/predicted L1 beam level measurement. |

If the answer to the above question is yes, one more question is that for RRM sub case 1, whether L1 RSRP difference between predicted and actual measurement needs to be also reported? Rapporteur believes the difference between RRM use case 1 and the other two sub cases are simulation methodology issue because eventually RAN2 is pursuing the prediction accuracy of L3 cell level measurement. In light of this, it may be not necessary to mandate company to report such L1 RSRP difference.

**Question 2.2.1-2: For RRM sub case 1, which option do you prefer?**

**Option 1: No L1 RSRP difference is necessary**

**Option 2: L1 RSRP difference is reported optionally i.e., up to company**

**Option 3: L1 RSRP difference is reported as mandatory**

|  |  |  |
| --- | --- | --- |
| Company | Preferred option | comments |
| NTT DOCOMO | Option 3 or Option 2 | Option 3 is more preferred. |
| OPPO | Option 2 | Once we aligned main metric, it is not so important to report L1 RSRP difference. but we are fine for company to report together with simulation result. |
| Apple | Option 2  (also see comments) | To us sub case 1 itself is optional and as we mentioned above we eventually don’t need all three, so maybe we can start the down-selection by eliminating sub case 1? |
| Mediatek | Option 3 | In case 1, the L1 RSRP difference is the definitive KPI for evaluating the AI/ML model's performance. Additionally, this metric can help us determine the applicability and extent to which L1 predictions in AI-BM are relevant for AI-driven mobility decisions. |
| Huawei, HiSilicon | Option 1 for L1 beam results | Agree with the rapporteur, there is no need to show L1 beam level results. The output should be L3 measurements. On the other hand, we should also evaluate the accuracy of L3 beam level predictions. |
| Samsung | Option 1 (or Option 2) | Agree with the rapporteur. We do not need to mandate L1 results for this study whose scope is for L3 mobility. |
| vivo | Option 3 | Since companies may process the L1 RSRP with different L1 or L3 filter parameters, we think it is needed to report L1 RSRP difference to show the AI model performance without filtering. Besides, it does not bring additional effort for simulation because the L3 RSRP is also derived from the L1 RSRP. |
| Ericsson | Option 1 |  |
| Xiaomi | Option 1 | The final target is cell level measurement result. L1 beam RSRP difference may not be able to reflect the L3 cell RSRP difference, since L1 beam selection and filtering is needed to generate L3 cell RSRP. The accuracy should also based on L3 cell RSRP difference. |
| CMCC | Option 3 | Same view with Q2.2.1-1. |
| ZTE | Option 1  Option 2 is also acceptable | We think the question is about L1 **beam** level RSRP difference.  L1 beam level RSRP difference is feasible for RRM sub case 1. Company can show the results if they want. But if we intend to compare sub case 1~3, then only L3 cell level RSRP difference can be used. |
| Nokia | Option 2, 3 |  |
| Intel | Option 2 |  |

There are mainly 3 ways to express RSRP difference [20][21][22][17]:

Option 1: CDF of RSRP difference

Option 2: Average RSRP difference

Option 3: RMSE of RSRP difference

Option 4: X dB margin of RSRP difference

Option 1 is basically a curve which record the RSRP difference of the whole prediction process. It is helpful to reflect the detail performance with clear picture but maybe difficult to capture in the TR.

Option 2 and option 3 are both a value reflecting the whole process. Mathematically RMSE can reflect the range of the RSRP difference better. Option 4 is a percentage of the RSRP difference, which is lower than X db. In some case e.g., when a threshold is needed in [Question 2.2.2-2](#OLE_LINK2), a value or percentage is more useful than a curve. But it seems also a bit redundant if all of them are used.

**Question 2.2.1-3: Among listed 4 options, which one(s) do you prefer to be taken as metric of RRM measurement prediction use case? If you have more option to add, please provide your description.**

|  |  |  |
| --- | --- | --- |
| Company | Preferred option(s) | comments |
| NTT DOCOMO | Option 2 and Option 1 | At least Option 2 can be mandatory. |
| OPPO | Option 3 and option 4 | For option 4, we suggest X is in range of {1,2,3}db |
| Apple | Option 2 or 3 | We hope a single option is selected – the more reporting options we have, the harder it is to evaluate the results. |
| Mediatek | Option 1, 2, 3 | Option 2 is mandatory.  Option 1,3 are optional. |
| Huawei, HiSilicon | Option 1 and 3 | We support option 1 as we think it is good idea to provide “a curve which record the RSRP difference of the whole prediction process”, but probably this is not really a CDF?  RMSE is on the other hand a metric that can be compared easily and gives overall idea of the accuracy of the predictions in the simulations. |
| Samsung | Option 2 | Option 2 can be a baseline. |
| vivo | Option 1 and Option 3 | RMSE of RSRP difference should be mandatory provided to show the overall RSRP prediction performance. And the CDF curve of RSRP difference can be optionally provided to show whether the AI model performs well for some samples and poorly for other samples. |
| Ericsson | Option 1 | Option 1 provides more information than other options. |
| Xiaomi | Option 2 or 3 | Difficult to define X dB in option 4. |
| CMCC | Option 1, Option 2 | We prefer to take Option 1 as the baseline of RSRP difference. |
| ZTE | Option 1 | Compared with other options, Option 1 can provide more information, and Option 1 can directly show how bad the worst prediction results could be. |
| Nokia | Option 1 (most important), 2, 3 | Overall, either of these options are easy to compute if/when we have the desired outcome. We think Option 1 is most informative. Options 2 and 3 can also be considered if/when needed. |
| Intel | Option 2 as baseline | other options are not precluded if companies want to bring simulation results |

RAN2 agreed that “measurement reduction rate as one KPI”, however, there is no detailed definition. For intra-cell prediction the definition could be different between temporal domain and spatial domain prediction. For temporal domain RAN2 start with pure temporal domain for case A and case B. Here is rapporteur’s recommendation of the definition:

Measurement reduction rate in temporal domain (MRRT):

MRRT= skipped measurement time instances / total measurement time instances

Measurement reduction rate in spatial domain (MRRS):

MRRS = skipped beams to be measured/ total beams to be measured

**Question 2.2.1-4: Do you agree the recommended definition of MRRT and MRRS? If you have different opinions, please provide your recommendation**

|  |  |  |
| --- | --- | --- |
| Company | Position: yes, no | comments |
| NTT DOCOMO | Yes (w/ comments) | Similarly, the KPI for frequency domain prediction can be defined as the ratio of the reduced measurement gap. |
| OPPO | Yes |  |
| Apple | Yes |  |
| Mediatek | Yes |  |
| Huawei, HiSilicon | Yes, but see comments | For MRRT, it would be worth clarifying that we assume all measurement time instances have the same length, otherwise, it will not be correct. |
| Samsung | Yes |  |
| vivo | Yes |  |
| Ericsson | Yes |  |
| Xiaomi | Yes |  |
| CMCC | Yes (see comments) | For inter-cell spatial domain measurement predictions (cell level measurements), MRRS = skipped cells to be measured/ total cells to be measured. For frequency domain prediction, the measurement gaps reduction shoule be defined. |
| ZTE | Yes for intra-cell case | Agree these two can be considered for “intra-cell prediction”. Should we also discuss the KPI for inter-cell? e.g. skipped cells? (as CMCC commented). For inter-freq case, besides gap reduction, we can also consider “skipped freqs”.  Note: For inter-freq measurements, even if the measured frequencies is reduced from 3 freqs to 1 freq and measurement gap is unavoid, there is also benefit since the measurement periodicity of each frequency can be shorten (MR can be triggered timely). |
| Nokia | Yes |  |
| Intel | Yes |  |

### Methodology

The cell level results refer to L3 cell level measurement results, which is depicted as reference point C in the measurement model. But it is not clear what does “beam level results” mean in the agreements. It could be the raw L1 beam level measurement result without L1 filtering (depicted by reference point A) or L1 beam level measurement result after L1 filtering (depicted by reference point A1). Because L1 filtering is up to UE’s implementation, so it is not easy to explain the difference between reference A and A1. Note this issue is also related to question 2.3.1.5-1 i.e., whether fast fading should be modelled as part of the channel modelling. Without fast fading element, there is not much difference between reference point A and A1. With fast fading element, L1 raw data before L1 filtering can reflect the channel variation better.

**Question 2.2.2-0: For the “beam level results” in RRM case 1 and RRM case 3, which option do you prefer?**

Option1: It is raw L1 beam level measurement result without L1 filtering i.e., reference point A

Option2: it is L1 beam level measurement result after L1 filtering i.e., reference point A1

Option3: it is up to company to choose either reference point A or point A1 and report it when providing simulation result

|  |  |  |
| --- | --- | --- |
| Company | Preferred option | comments |
| NTT DOCOMO | Option 3 | We think this is too detailed and should be up to the companies. |
| OPPO | Option 1 | This option help to figure out real measurement reduction in relatively easy way. |
| Apple | Option 3 | Agree with DoCoMo. |
| Mediatek | Option 3 | Agree with Docomo. Given that AI/ML models are based on implementation and will not be standardized, companies should have the freedom to select their inputs. However, for evaluation purposes, it would be beneficial if companies could disclose which options they have employed in their reporting. |
| Huawei, HiSilicon | Does not matter | This does not really matter as we are only interested in the output which is cell level measurement. Companies may provide additional information about what input they used for their model. As mentioned before, what we need to align though is how beam consolidation/selection is performed to obtain cell level result based on beam level measurement. |
| Samsung | Option 2  (Ok with Option 3) | Although L1 filtering is UE implementation, L1 filtering is essential to eliminate the fast fading effect.  We are not sure how raw L1 beam measurement can be helpful to figure out real measurement reduction as real measurement should use L1 filtering by implementation. |
| vivo | Option 3 |  |
| Ericsson | Option 3 |  |
| Xiaomi | Option 2 for NW side model  Option 3 for UE side model | For UE side model, the input can be up to UE implementation.  For NW side mode, UE need to report the input to NW. L1 measurement result after L1 filtering is used in RAN BM prediction, so RAN2 can reuse this assumption. |
| CMCC | Option 3 |  |
| ZTE | Option 2 | Agree with Samsung, L1 filtering is essential to eliminate the fast fading effect. Besides, option 1 would consume lots of computing resources, since the inference is too frequent. |
| Nokia | Option 3 | We think UEs should report same as in legacy, with the difference now that (depending on the configuration), the reports can contain either measurements or predictions. |
| Intel | Option 3 | Agree with HW that L3 filtering parameters are more important in sub case 1. |

RAN2 agreed at RAN2#125bis:

1. We will consider intra-frequency intra and inter-cell spatial domain measurement predictions, for beam and cell level measurements.
2. For temporal domain measurement prediction, we will consider the AI-PHY beam management Case A and Case B from the RAN1 AI/ML PHY TR and it applies to both beam level and cell level. As baseline we will focus on pure temporal prediction.

These two agreements imply RRM prediction could be categorised in following 3 dimensions:

**D1: intra-frequency or inter-frequency.**

As agreed during last meeting, there are 3 cases totally i.e., FR1\_to\_FR1 intra-frequency, FR1\_to\_FR1 inter-frequency and FR2\_to\_FR2 intra-frequency.

**D2: intra-cell or inter-cell or cluster approach.**

The intra-cell prediction basically means the input and output measurement result of the model comes from same cell. Obviously intra-cell can only be applied for intra-frequency case. Inter-cell prediction means the input and output of the model is different cell. It could be for either intra-frequency or inter-frequency case. It is not crystal clear what is cluster approach. Based on some offline discussion with proponents, it basically means the number of input cells could be more than one cell [18] for intra-frequency prediction.

**D3: Temporal or spatial domain**

For temporal domain, RAN2 agreed that we will mimic case A and case B of BM case 2 in [38.843] in pure time domain as baseline. As for spatial domain prediction, it basically means UE will measure partially configured beams instead full set beams to perform RRM measurement. For RRM sub case 1, it also means that beam level measurement result of partial beams (i.e., not measured ones) is predicted based on measurement of other beams (i.e., measured ones).

The 1st step we can do is to list all the potential combinations and check which of them are valid case to be discussed.

|  |  |  |
| --- | --- | --- |
| Scenarios | FR1\_to\_FR1 or FR2\_to\_FR2 intra-frequency | FR1\_to\_FR1 inter-frequency |
| Intra-cell | Intra\_F\_C\_T: temporal domain, to be clarified  Intra\_F\_C\_S: spatial domain (FR2\_to\_FR2 only), to be clarified | Invalid case |
| Inter-cell | Intra\_F\_Inter\_C: To be clarified | Inter\_F\_C: to be clarified |
| Cluster approach | Intra\_F\_Cluster: To be clarified | Invalid case |

Table 2.2.2-2 prediction combinations

**Question 2.2.2-1: Do you agree with listed combinations in Table 2.2.2-2? If you have more cases to be discussed, please provide your case with detail description.**

|  |  |  |
| --- | --- | --- |
| Company | yes or no | Comments or more case(s) |
| NTT DOCOMO  (updated) | No | The cluster approach is still valid for the FR1\_to\_FR1 inter-frequency case because the model's input can be the measurements of a cluster of cells on Band #1, and the model's output can be the predicted value for another cluster of cells on Band #2. Of course, the two clusters include different cells, but it is still a cluster-based approach.  Rapporteur: Because this is additional one, I assume you are at least fine with the combinations listed in the table, right? The cluster approach is addressed in question 2.2.2-13  DCM: Thank you for your reply. Yes, the existing combinations in the table are fine. Our concerns are about the two “Invalid case” in the table.  For all 4 cases about Inter-cell and cluster approaches, the sub-cases of T and S domains are also available. |
| OPPO | Yes |  |
| Apple | See comments | The way we understand the question, it merely proposes to use the terminology “intra/inter\_F\_intra/inter\_C\_T/S”. In this case it’s OK. |
| Mediatek | Yes, but.. | For RRM prediction, predictions in the temporal, spatial, and frequency domains are considered as the potential solution. We can allow companies the flexibility to explore the benefit of these solutions across various scenarios. At this stage, it might be unnecessary to explicitly define or limit the scenarios and cases.  We can categorize the scenarios/cases at a later time based on the interests and contributions of the companies.  Additionally, I agree with Docomo's observation that the cluster approach remains applicable to the inter-frequency scenario. |
| Huawei, HiSilicon | See comments | We think we can exclude the cluster approach at the moment. We think for now we should stick to what we agreed already, i.e.:   * FR1-to-FR1   + Focus on intra-frequncy in time domain prediction for the purpose of measurement reduction   + Study inter-frequency scenario in terms of which scenarios can be studied without requiring new channel model and also resolving any simulation assumptions (if possible). * FR2-to-FR2   + Focus on intra-frequency   + Perform evaluation both in time and spatial domain |
| Samsung | No | Agree with NTT DOCOMO. The cluster approach can be applied for inter-cell prediction where multi-frequency inputs can be used. |
| vivo | See comments | Fine with the intention.  However, prefer not to mix the temporal/spatial/frequency domain with the intra/inter- cell approach. The former ones are general output types while the latter ones are more about methodology. Besides, the definition of cluster approach and the difference between intra/inter- cell approach should be further clarified. |
| Ericsson | Yes, but | The Intra\_F\_C\_S case is not part of the SID, so it should not be considered. |
| Xiaomi | Comment | If cluster is in the scope, we also wonder why cluster is invalid in FR1\_to\_FR1 inter-frequency scenario. |
| CMCC | Yes (see comments) | We are fine with the combinations. But the cluster approach need to be clarified, and decide whether it is valid for FR1\_to\_FR1 inter-frequency, FR1\_to\_FR1 or FR2\_to\_FR2 intra-frequency. |
| ZTE | See comments | Agree with NTT DOCOMO, the cluster approach can be applied for inter-freq prediction. But considering the simulation complexity, we can consider it later. |
| Nokia | See comments | Agree with the intention and this can be a starting point. We think this needs to be further discussion in the context of sub-use-cases. |
| Intel | See comments | For simulation, even though we use agreements as basleine, further points still need to be clarified:   * FR1-to-FR1   + Focus on intra-frequncy in time domain prediction for the purpose of measurement reduction   (whether it considers both inter-cell and intra-cell? we think intra-cell should be evaluated with higher priority, that is **intra-freq intra-cell time domain**. Also for intra-cell, inference input also needs to be clarified: whether it’s single cell or same group of cells. To our understanding, **single cell** (i.e. serving cell or neighbouring cell only) is the baseline.)   * + Study inter-frequency scenario in terms of which scenarios can be studied without requiring new channel model and also resolving any simulation assumptions (if possible).   (in our understanding, this is **inter-freq inter-cell spatial domain**)   * FR2-to-FR2   + Focus on intra-frequency   + Perform evaluation both in time and spatial domain   (in our understanding, this is **intra-freq intra-cell time domain**, and **intra-freq intra-cell spatial domain**)  In summary, intra\_F\_Inter\_C and Intra\_F\_Cluster are not considered or low priority. |

Combination Intra\_F\_C\_T refers temporal domain for both FR1\_to\_FR1 or FR2\_to\_FR2 intra-frequency intra-cell prediction. RAN2 agreed to mimic case A and case B in TR [2] without mixing spatial domain as baseline. Here is the Figure for case A in [2]:



Figure 2.2.2-1

Case A basically mean measurement result in future of one cell e.g., cell A is predicted based on historical ones of the same cell A. It can be further illustrated with Figure 2.2.1-1:



Figure 2.2.2-2 Intra-cell temporal domain prediction – case A

The reason for UE to predict RRM measurement results in future is to report either the measurement results or other associated event e.g., measurement event to network in advance so that network can trigger handover in the right time. Rapporteur’s understanding is that such evaluation is targeting 2nd goal discussed before and hence no measurement reduction is necessary. Observation window refer to a duration UE perform the actual measurements. When UE perform measurement in predicted instance(s), that instance(s) becomes part of the observation window instead of prediction window as illustrated in Figure 2.2.2-2 i.e., observation window and prediction window will slide when more measurement(s) is performed by UE in temporal domain. The prediction window depends on inner elements like model performance and observation window length and also outer element like radio channel. Regardless of these elements, the predicted measurements within prediction window should meet some predefined prediction accuracy because otherwise it doesn’t make sense.

One example of description of methodology of Intra\_F\_C\_T\_Case A: Intra-cell temporal domain prediction is done by predicting measurement result(s) in prediction window based on measurement results in observation window of the same cell for both FR1\_to\_FR1 and FR2\_to\_FR2 intra-frequency scenario, where the prediction accuracy of the measurement result(s) in prediction window should be higher than one predefined threshold. The predefined threshold should be aligned among companies. The detail value is FFS.

**Question 2.2.2-2: How do you think of the example methodology** **of Intra\_F\_C\_T\_Case A? If have better formulation, please provide your recommendation.**

|  |  |  |
| --- | --- | --- |
| Company | comment | other formulation |
| OPPO | Agree |  |
| Apple | Not necessarily | We would be OK with just the first part “Intra-cell temporal domain prediction is done by predicting measurement result(s) in prediction window based on measurement results in observation window of the same cell for both FR1\_to\_FR1 and FR2\_to\_FR2 intra-frequency scenario”, but:   * What’s the point of this case, other than measurement event prediction? Measurement event prediction part of this study is supposed to be on hold for now. * Even if/when we get to the measurement event prediction, why do we need “accuracy … be higher than predefined threshold” as opposed to just evaluating accuracy? |
| Mediatek | Need more discussion.  Similar to AI-BM, both cases A and B could contribute to RS reduction. It is premature to conclude that case A specifically aims to improve HO performance without having conducted any evaluations. While it is possible that this speculation may hold true, a thorough assessment is necessary before confirming such a conclusion. | If the 1st goal is considered, the system performance in terms of HOF, RLF, Pingpong, ToS, data interruption time needs to be evaluated, as well as the tradeoff between prediction accuracy and the AI model complexity.  If the 2nd goal is considered, an alternative formulation can be evaluating the prediction accuracy alongside the tolerable degradation levels for the aforementioned KPIs. |
| Huawei, HiSilicon | It seems OK in general, but we think we can compare prediction accuracy vs. different lengths of prediction window, so we should agree on several values of prediction window. Not sure why we need to agree on the accuracy threshold at this stage. We do not know how prediction accuracy impact HO performance yet, so it is hard to come up with the good number. |  |
| Samsung | We do not see the prediction accuracy/threshold is necssary in the methodology. It is more like a condition or KPI for the prediction, not definition of use case. | This methodology can be replaced by RSRP difference, discussed in Q2.2.1-3 |
| vivo | Remove the second part related to accuracy and threshold. | From our understanding, it’s quite difficult to reach a consensus on the threshold of the prediction accuracy, as the final system-level performance relies on how to utilize the RRM prediction. |
| Ericsson | The prediction accuracy threshold is not necessary for the evaluation. The result is just reported. | Intra-cell temporal domain prediction is done by predicting measurement result(s) in prediction window based on measurement results in observation window of the same cell for both FR1\_to\_FR1 and FR2\_to\_FR2 intra-frequency scenario. |
| Xiaomi | We think the prediction accuracy is part of the performance evaluation. And the accuracy may be related to AI model and NW condition. How to ensure the accuracy should be done by performance monitoring. We don’t see the need to define a predefined accuracy threshold. |  |
| CMCC | We have concern whether the same predefined prediction accuracy could guarantee the same handover performance among the different platform of different companies, which may be related to company’s implementation. As we know, there is no predefined threshold in the evaluation of AI-PHY beam management. |  |
| ZTE | We think there is no need to define the prediction accuracy threshold. At this stage, we can focus on RSRP difference between actual results and prediction results at different future time instance. |  |
| Nokia | See comments | Is the intention to say that the UE would do some testing of the model on its own during the observation window, and once it observes that the predicted measurements do not deviate much from the true measurements (e.g., accuracy criterion is satisfied), UE switches to only making predictions?  Overall, we think it would be simpler to consider fixed length of the observation/prediction window as starting point. Then we can discuss how to make it more dynamic based on accuracy or other metrics. In our view, accuracy alone is not sufficient to determine the switch between observation and prediction windows. Overall, we think this dynamic switch might also have some implications on the LCM-level analysis. |
| Intel | Agree with above companies that the accuracy threshold cannot be predefined without any simutlation evaluation. |  |

For temporal domain prediction case B, here is the Figure in [2]:



Figure 2.2.2-3

In [2] ‘s description of case B: “based on a periodicity T of the required reference signals for measurements to achieve a certain beam prediction accuracy. An example is shown in Figure 6.3.1-3.

- For non-AI baseline (Option 2), every T=X ms reference signals for measurements are needed

- For AI, every T=Y ms, reference signals for measurements are needed”

Here non-AI baseline (option 2) is “sample and hold based on the previous measurements” in [2]. For RAN2’s evaluation, to make it simple no such baseline is needed i.e., in the benchmark case all measurement instances are measured by UE. So essentially case B means some of the time instances are skipped by UE, whose measurement results will be predicted based on measured instances as illustrated in Figure 2.2.2-4.



Figure 2.2.2-4 intra-cell temporal domain prediction – case B

In order to compare among companies, it seems necessary to align the measurement reduction rate so that prediction accuracy can be compared with each other. For the same reduction rate, the skipping pattern i.e., which instances are skipped and hence predicted in temporal domain could be left to company’s implementation because otherwise there are too much combinations. Rapporteur’s view is that the result is comparable as long as reduction rate is aligned and believe detail pattern doesn’t matter too much. Obvious case B is targeting goal 1.

Example methodology of Intra\_F\_C\_T\_Case B: Intra-cell temporal domain prediction is done by predicting sub set measurement instances in temporal domain of the same cell for both FR1\_to\_FR1 and FR2\_to\_FR2 intra-frequency scenario. The measurement reduction rate should be aligned among companies. The detail value is FFS.

**Question 2.2.2-3: How do you think of example methodology of Intra\_F\_C\_T\_Case B? If you have better formulation, please provide detail description.**

|  |  |  |
| --- | --- | --- |
| Company | comment | other formulation |
| OPPO | agree |  |
| Apple | Agree (with comments) | We probably need multiple reduction rate values. |
| Mediatek | Agree  Need to consider both measurement reduction rate and the prediction accuracy as intermediate KPI to evaluate the AI model performance. | Considering the necessity to assess system performance through KPIs such as HOF, RLF, Pingpong, ToS and data interruption time, an alternative formulation can be evaluating the RSRP difference alongside the tolerable degradation levels for the aforementioned KPIs. |
| Huawei, HiSilicon | Agree with the comment from Apple. We think the goal is to compare accuracy against vs measurement reduction rate, so several values for measurement reduction rate should be agreed for which the companies should provide their results. |  |
| Samsung | Agree |  |
| vivo | Agree with comments | 1. Remove the description related to reduction rate.  2. We understand case B intends to extend the periodicity of the measurement and utilize the prediction to replace the instances that were originally required for measurement. Since Case B also has concepts of observation window and prediction window, the current definition of case B can be refined to further distinguish it from case A:  Intra-cell temporal domain prediction is done by extending the measurement periodicity and utilizing the prediction to represent the instances that should be measured in temporal domain of the same cell for both FR1\_to\_FR1 and FR2\_to\_FR2 intra-frequency scenario. |
| Ericsson | Agree. |  |
| Xiaomi | Agree |  |
| CMCC | Agree |  |
| ZTE | Agree with Apple and Huawei. Multiple reduction rate values are needed. |  |
| Nokia | Agree, but see comments | Agree with the intention. However, we think that additional intermediate KPIs that capture the trade-off between the saved measurements and the loss in accuracy need to be considered to ensure that the results are reliable. |
| Intel | Agree | Agree with Apple that multiple reduction rate needs to be considered. |

For both Intra\_F\_C\_T\_Case A and Intra\_F\_C\_T Case B, rapporteur’s understanding is that they are both applicable for 3 RRM sub cases. One more common issue is whether sampling period should be aligned among company or not. Note usual RAN4 assume UE will oversample within one measurement period. The minimum measurement period for FR1 and FR2 intra-frequency measurement is different. It hints the sampling period between FR1 and FR2 could be also different, if necessary.

**Question 2.2.2-4: For both Intra\_F\_C\_T\_Case A and Intra\_F\_C\_T Case B, which RRM sub cases are applicable?**

|  |  |  |
| --- | --- | --- |
| Company | applicable RRM sub cases | comments |
| NTT DOCOMO | All 3 sub cases | For Intra\_F\_C\_T\_Case A. We would like to clarify the meaning of CB\_1\_1\_Case B  Rapporteur: sorry this is a typo. it should be intra-frequency intra-cell time domain case B i.e. Intra\_F\_C\_T Case B |
| OPPO | All sub cases |  |
| Apple | 2 and 3 | We are not sure 1 adds much to the study (other than unnecessary complexity) |
| Mediatek | All 3 sub cases | Mediatek |
| Huawei, HiSilicon |  | This does not really matter, we do not have to spend more time on applicability of RRM sub cases. This is AIML model implementation issue and companies can simply indicate which method they used in their model. |
| Samsung | All sub cases |  |
| vivo | All sub cases |  |
| Ericsson | All 3 sub cases |  |
| Xiaomi | All | Theriotically, all are applicable. But we are also fine to exclude use case 1 for simplicity. |
| CMCC | All 3 RRM sub cases |  |
| ZTE | All 3 RRM sub cases |  |
| Nokia | All |  |
| Intel | All sub cases | Agree with HW that it is up to companies’s implementation which sub case is considered. |

**Question 2.2.2-5: For both Intra\_F\_C\_T\_Case A and Intra\_F\_C\_T Case B, do you think it is necessary to align sampling period? If so, please recommend sample period for both FR1 and FR2 respectively.**

|  |  |  |
| --- | --- | --- |
| Company | Position: yes or no | Comments |
| NTT DOCOMO | Yes | Align with the typical SSB burst periodicity, i.e., 20ms. |
| OPPO | Yes | Our suggestion is 40 ms which is captured in 36.839. |
| Apple | Not necessarily | We don’t see how using different sampling rates in different evaluations by different companies would be a problem. Remember, we are evaluating AI/ML algorithms, not writing UE RAN4 requirement.  Consider Measurement Reduction Rate or accuracy metrics – even if different companies use different sampling rates, the results of such AI/ML models would still be consistent and comparable. |
| Mediatek |  | The sampling period should be within a specified range; however, it is unclear if alignment to an exact value is necessary. Should an exact value be required, a period of 20ms is preferred. |
| Huawei, HiSilicon | Yes | If we do not align sampling period for measurements, then it may be hard to truly compare the results and hard to draw conclusions from them. |
| Samsung | Yes | No strong view on sample period |
| vivo | Yes | 20ms can be used as baseline. |
| Ericsson | Yes | The sampling period can be the SSB burst period or a multiple of it. |
| Xiaomi | Yes |  |
| CMCC | Yes | Same view with OPPO. |
| ZTE | Yes | Alignment of sampling period is helpful for comparing the simulation results among companies. We also think 20ms can be considered as a baseline for both FR1 and FR2.  In addition, we think the periodicity of L1 filtering also needs to be discussed. e.g. how many L1 simples will be used to generate one L1 level result. |
| Nokia | Yes | We can consider 20ms as baseline. Other values are also not precluded. |
| Intel | Yes |  |

For Intra\_F\_C\_S spatial domain prediction, the way to do it is different among RRM sub cases. For RRM sub case 1, it basically means the L1 beam level measurement result of sub set of configured reference signal e.g. SSB is predicted based on L1 beam level measurement result of other RS. And then L3 cell level measurement is got by UE after post processing, namely consolidation and L3 filtering. This is illustrated in Figure 2.2.2-5



Figure 2.2.2-5 Intra\_F\_C\_S intra-cell spatial domain prediction for RRM sub case 1

For RRM sub case 3 there is no such consolidation and L3 filtering. The procedure can be illustrated in Figure 2.2.2-6:



Figure 2.2.2-6 Intra\_F\_C\_S intra-cell spatial domain prediction for RRM sub case 3

In order to compare between the RRM sub case 1 and 3 and also to compare among companies result, it is necessary to align measurement reduction rate e.g., how much percentage of measurement of the SSB beams is skipped. Then by comparing the same metrics i.e., RSRP difference, we can know which method is better. If this is agreeable, one further issue is that whether company also need align detail skipping pattern in spatial domain. Rapporteur’s view is that this could be simply left to company because there could be so much detail pattern which doesn’t make too much difference with each other.

For RRM sub case 2 i.e., L3 to L3, it is not clear how to do it in spatial considering the input measurement is already L3 cell level measurement result. So, it seems Intra\_F\_C\_S is not applicable for RRM sub case 2.

Example methodology of Intra\_F\_C\_S: Intra-cell spatial domain prediction is done by measuring sub set of configured SSB as input to the model to predict L3 cell level measurements for every instance of the same cell. It is only applicable for FR2 intra-frequency scenario and RRM sub case 1 and 3. The measurement reduction rate should be aligned among company without defining detail pattern. The detail rate value is FFS.

**Question 2.2.2-6: How do you think of example methodology of Intra\_F\_C\_S? If you have better formulation, please provide detail description.**

|  |  |  |
| --- | --- | --- |
| Company | comment | other formulation |
| OPPO | Agree |  |
| Apple | Agree (with comments) | We may need multiple agreed reduction rates. |
| Mediatek | It's uncertain if this applies exclusively to FR2. From what I understand, there is also multi-beam operation in FR1, albeit with a fewer number of beams compared to FR2. | Considering the necessity to assess system performance through KPIs such as HOF, RLF, Pingpong, ToS and data interruption time, an alternative formulation can be evaluating the RSRP difference alongside the tolerable degradation levels for the aforementioned KPIs. |
| Huawei, HiSilicon | The definition seems OK in general, but we should agree on multiple values of measurement reduction rate to see how it impacts prediction accuracy. We also should not say that it is only applicable to FR2 as it may be applicable to FR1 as well. We just agreed not to evaluate it for FR1 which is a different thing. |  |
| Samsung | Agree |  |
| vivo | OK as baseline for Cell-level | The definition for beam-level may refer to the BM-Case1, i.e., Spatial-domain Downlink beam prediction for Set A of beams based on measurement results of Set B of beams.  Besides, prefer to decouple the definition from the performance KPI (measurement reduction rate) |
| Ericsson | If we agree on this scenario, the formulation is OK. |  |
| Xiaomi | Agree |  |
| CMCC | Agree |  |
| ZTE | Agree with Huawei. At the early stage of SI, RAN2 agreed to not evaluate the spatial domain measurement prediction in the FR1 scenario, but this does not mean spatial domain prediction is not applicable in FR1. |  |
| Nokia | Agree | Agree that the trade-off between reduction rate and accuracy should be evaluated. As an example, different reduction rate values can be considered, and companies can report the corresponding accuracies. |
| Intel | Agree |  |

For Intra\_F\_Inter\_C and Inter\_F\_C, one relevant agreement is “We will consider intra-frequency intra and inter-cell spatial domain measurement predictions, for beam and cell level measurements”. If spatial domain measurement prediction is also applicable for inter-cell case, it basically means UE predict measurement of another cell (let’s say cell B) by measuring only **partial SSB** of the one cell (let’s say cell A). The intention of inter-cell prediction is to save measurement overhead of another cell. If it could be done the measurement reduction rate is 50% technically which is already plausible. Further reduction in cell A could make the prediction performance much worse. During offline discussion with proponent rapporteur’s view is confirmed. So for both Intra\_F\_Inter\_C and Inter\_F\_C, the measurement of cell A should not be reduced in both temporal and spatial domain.



Figure 2.2.2-7 inter-cell prediction

**Question 2.2.2-7: For both Intra\_F\_Inter\_C and Inter\_F\_C, do you agree that the measurement on source cell (cell A in the example) should not be reduced in both temporal and spatial domain? If no, please clarify which domain(s) can be reduced and why.**

|  |  |  |
| --- | --- | --- |
| Company | Position: yes or no | comments |
| NTT DOCOMO | Yes | At least for the baseline case, we start from the non-reduced measurements. Measurement reduction on the source cell can be further studied. |
| OPPO | Yes |  |
| Apple | Yes |  |
| Mediatek | Yes | Agree with Docomo. |
| Huawei, HiSilicon | Yes |  |
| Samsung | Yes | Fine to start with a simple assumption |
| vivo | Yes | It will be difficult to reduce measurements of both source cell and neighbour cell(s) at the same time, so we should first focus on neighbour cell(s) measurement prediction with full source cell measurement results. |
| Ericsson | Yes | We think it is not needed to reduce furtherly measurements on source cell. |
| Xiaomi | Yes | Agree this can be baseline. But we can further study whether further measurement reduction can be done in next step. |
| CMCC | Yes | Same view with NTT DOCOMO |
| ZTE | Yes |  |
| Nokia | Yes |  |
| Intel | Yes |  |

For Inter\_F\_C i.e., inter-frequency and inter-cell prediction for FR1\_to\_FR1 case, one task is left as following:

**Agreements to start evaluations**

* FR1-to-FR1
  + Focus on intra-frequncy in time domain prediction for the purpose of measurement reduction
  + Study inter-frequency scenario in terms of which scenarios can be studied without requiring new channel model and also resolving any simulation assumptions (if possible).
* FR2-to-FR2
  + Focus on intra-frequency
  + Perform evaluation both in time and spatial domain

The relevant section of 38.901[19] is section 7.6.5. Rapporteur’s understanding is that existing channel modelling in [19] only cover co-located scenario i.e., the discussion on channel modelling need be open by RAN1 for non-co-located case. And during offline discussion with proponents, it is also confirmed that some operators and vendors are fine to start from co-located. One thing needs to clarify that co-located scenario doesn’t mean prediction can only be done in the site where serving cell is located. For neighbouring site UE can also predict neighbouring cell on non-serving frequency by measurement neighbouring cell in serving frequency.

**Question 2.2.2-8: For Inter\_F\_C, do you agree RAN2 start evaluation from co-located scenario? If no, please clarify what scenario is necessary.**

|  |  |  |
| --- | --- | --- |
| Company | Position: yes or no | comments |
| NTT DOCOMO | Yes |  |
| OPPO | Yes |  |
| Apple | Yes |  |
| Mediatek | Yes |  |
| Huawei, HiSilicon | Yes | We agree with the proposal from the rapporteur. For non-collocated case, it does not seem to be possible to do it without channel modelling work which we think should be avoided. |
| Samsung | Yes | Agree with the rapporteur that TR 38.901 model covers only co-location scenario. |
| vivo | Yes | This will be easier for us to start the inter-frequency evaluation. |
| Ericsson | Yes |  |
| Xiaomi | Yes |  |
| CMCC | Yes | Non-co-located case could be studied later. |
| ZTE | Yes |  |
| Nokia | Yes |  |
| Intel | Yes |  |

If question 2.2.2-8 is confirmed, further question is what is the relationship between source cell (say cell A) and target cell (say cell B)? There are could be two cases:

* Case 1: cell A and cell B is in the same sector (assuming there are 3 sectors per gNB site)
* Case 2: cell A and cell B is neighbouring sector in the same gNB site

 

Figure 2.2.2-8.1 case 2 Figure 2.2.2-8.1 case 1

After offline discussion with company, rapporteur believe inter-frequency prediction itself is already difficult compared to other cases. The inter-sector prediction will make the situation even worse. In order to have reasonable prediction accuracy, RAN2 should focus on case 2.

**Question 2.2.2-9: Do you agree for Inter\_F\_C, RAN2 should focus on the case where cell A and cell B are in the same sector? If no, please clarify reason to support case 1 or other case.**

|  |  |  |
| --- | --- | --- |
| Company | Position: yes or no | comments |
| NTT DOCOMO | Yes(need clarification) | It seems that the above descriptions of Case 1 and Case 2 are not aligned with the figures and the texts following the figures. We support that Cell A and Cell B are in the same sector.  Rapporteur: thanks for spotting this😊 |
| OPPO | Yes |  |
| Apple | Yes |  |
| Mediatek |  | We can use this case as starting point. But we can also explore the possibility that cell A and cell B are not exactly overlapped later. |
| Huawei, HiSilicon | Yes | For the different sector case, we still have the channel modelling issue as it is not truly a collocated scenario. |
| Samsung | Yes |  |
| vivo | Yes | This will be easier for us to start the inter-frequency evaluation. |
| Ericsson | Yes |  |
| Xiaomi | Yes |  |
| CMCC | Yes |  |
| ZTE | Yes | We can start from the case that cell A and cell B are in the same sector. If time allows, we can study other cases later., |
| Nokia |  | Agree with Mediatek. The input measurements can and often have to be from more than one cell to learn the correlation. This can be a starting point; however non-overlapping cells A and B should also be kept in scope. |
| Intel | Yes |  |

For Inter\_F\_C it is not clear which RRM sub cases are applicable. RRM sub case 1 means beam level prediction is feasible between cell A and cell B one by one. RRM sub case 3 and case 2 cover the detail and imply L3 cell level measurement of cell B can be predicted based on either L1 beam level or L3 cell level measurements of cell A. They look more promising from rapporteur point of view.

**Question 2.2.2-10: Among RRM sub case 1,2,3, which one(s) is applicable for Inter\_F\_C?**

|  |  |  |
| --- | --- | --- |
| Company | Position: applicable RRM sub case(s) | comments |
| NTT DOCOMO | Alll 3 sub cases. | In our view, there are not enough results to support the down selection so far. |
| OPPO | RRM sub case 2 or 3 and slightly prefer 2. | It is not clear what does it mean for sub case 1. Does it mean the index of the target beam in cell B should be the same as in cell A, or it is kind of cluster approach in beam level? It implies the beam related configuration between cell A and cell B should be strictly aligned with each other. For sub case 2 and 3 there is no such restriction. Plus this is for FR1 to FR1 scenario, where there is only few beams or no beams at all. |
| Apple | 2 is sufficient |  |
| Mediatek | All 3 sub cases. | For FR1, we may still have multiple beam operation, although it has fewer beams than FR2. |
| Huawei, HiSilicon | Does not matter | Companies are free to use the model they see fit and can simply clarify when providing the results what has been used as input. |
| Samsung | Case 2 | Inter-frequency scenario is for only cell-level prediction. We agree with OPPO that Case 1 does not make sense. Considering inter-frequency evaluation only for FR1, Case 2 looks ok. |
| vivo | All 3 sub cases. |  |
| Ericsson | All 3 sub cases. |  |
| Xiaomi | All | Theriotically, all are applicable. But we are also fine to exclude use case 1 for simplicity. |
| CMCC | All 3 RRM sub cases | We need evaluation results to verify which RRM sub cases are applicable for Inter\_F\_C. |
| ZTE | All 3 RRM sub cases | We think there is no restriction for the beam index in cell A and cell B if the model training and data collection are performed per site. |
| Nokia | Case 2 as starting point, other sub cases not precluded | It should be up to companies to decide how they design their dataset and models. What matters is that the documentation they provide is clear and complete. |
| Intel | all three use cases |  |

When UE perform intra-frequency measurement, UE’s behaviour could be bit different between serving cell and neighbouring cell. But there is no difference between neighbouring cell regardless whether neighbouring cell is co-located with serving cell or not. But it is almost not feasible for UE to predict a non-co-located neighbouring cell by measuring serving cell since there is spatial consistency between them. So technically it may make sense to predict a co-located neighbouring cell by measuring serving cell.



Figure 2.2.2-9

**Question 2.2.2-11: Do you agree for Intra\_F\_Inter\_C, the main case is to predict co-located neighbouring cell by measuring serving cell? If no, please provide description of other case(s).**

|  |  |  |
| --- | --- | --- |
| Company | Position: yes or no | comments |
| NTT DOCOMO | No | Because the HO does not always occur among co-located neighbouring cells, we should consider more general cases. Besides, AI/ML should be stronger than legacy algorithms on complex tasks such as non-co-located cases. |
| OPPO | Yes | Disagree with NTT. First of all, we think it is not necessarily due to nature of intra-frequency measurement. Secondly for non-co-located cells we even don’t understand how could it be feasible i.e. there is no such corelation among them. |
| Apple | Yes | This is a reasonable simplification |
| Mediatek | No | We have some sympathy on Docomo’s view. It’s too early to draw the conclusion without discussion and evaluation.  It depends on what kind of corelation we are talking about. Theoretically and in actual networks, the multiple-path components from adjacent cells to the same UE, although different, may have correlations due to being related to similar interacting objects and environment. However, with the current stochastic models, this is not observable. But such corelation can be observed in the deterministic channel model. |
| Huawei, HiSilicon | No | The UE cannot distinguish whether the beams are from the same gNB or not. |
| Samsung | No | Co-located inter-sector assumes fully correlated (aligned) path-loss shadow fading and large-scale parameter. So the prediction will be more accurate than non-co-location. However, the study needs to see the prediction performance for non-colocation without channel correlation. |
| vivo | No | In our view, prediction for non-co-located neighbouring cell by measuring serving cell is also possible and should also be considered. The channel characteristics may be less relevant for non-co-located cells compared with co-located cells, but with footprint-like scheme, AI/ML model can still find some relationship between the measurement results of non-co-located cells |
| Ericsson | No | We do not see the benefit of limiting the study at this point in time. |
| Xiaomi | No | Use case may need further discussion. If UE only measures serving cell, it may be difficult to predict neighour cell. However, if UE can measure a cluster of cells and the input includes measurements of multiple cells, including serving and neighbour cells. It’s possible for UE to predict other neighbour cells. |
| CMCC | No | It is too early to preclude non-co-located neighbouring cell prediction without evaluation results. |
| ZTE | No | Non-located neighbour cell prediction is also feasible. |
| Nokia | No | Agree we can start with co-located, but the prediction can be also not co-located cell and the input can be serving cell and a number of other cell from the serving cell frequency. |
| Intel | No | In our understanding, this is possible if footprint are considered, especially for NW-side model. |

Assuming the answer to question 2.2.2-10 is yes, the best case is that measurement of only two co-located neighbouring cells can be saved i.e., the gain is limited. Secondary when UE measure serving cell, the signal of other neighbouring cell including co-located ones will be most likely also received, in that case intra-frequency intra-cell can be conducted, which is technically more reliable. So, it seems not so attractive for RAN2 to evaluate Intra\_F\_Inter\_C.

**Question 2.2.2-12: Do you agree Intra\_F\_Inter\_C will not be evaluated at least in early stage?**

|  |  |  |
| --- | --- | --- |
| Company | Position: yes or no | comments |
| OPPO | Yes |  |
| Apple | Yes |  |
| Mediatek |  | It’s OK to evaluate this later. |
| Huawei, HiSilicon | Yes | This scenario does not seem to bring much benefits for measurement reduction as the cells are both on the same frequency so can be measured simultaneously. |
| Samsung | Yes | Start with simpler scenarios |
| vivo | Yes | Although Intra\_F\_Inter\_C prediction is possible from our analysis above, we admit that Intra\_F\_Intra\_C prediction is easier and should be studied first. |
| Ericsson | Yes |  |
| Xiaomi | Yes |  |
| CMCC | Yes with comments | We think Intra\_F\_Inter\_C is a general case which should be considered. |
| ZTE | Yes | We are fine to evaluate this scenario later. |
| Nokia | Yes | Ok to consider this at a later stage, once the simulation work is more consolidated. |
| Intel |  | ok to evaluate in later stage. |

It is not crystal clear what does cluster approach mean. So, before RAN2 has common understanding it is very difficult to proceed with this approach. Rapporteur believe following issues should be first clarified:

Issue Intra\_F\_Cluster\_1: Among FR1\_to\_FR1 intra-frequency, FR1\_to\_FR1 inter-frequency and FR2\_to\_FR2 intra-frequency, which scenario is applicable?

Issue Intra\_F\_Cluster\_2: Is it applicable for co-located cells or also non-co-located cells?

Issue Intra\_F\_Cluster\_3: Assuming the number of input and output cells is IN\_N and OUT\_N respective, what is the relationship between IN\_N and OUT\_N e.g. should OUT\_N<= IN\_N, can OUT\_N>1, what is maximum number of IN\_N and OUT\_N etc.?

Issue Intra\_F\_Cluster\_4: Any adjustment on metrics is needed for cluster approach?

**Question 2.2.2-13: Company are kindly requested to answer Intra\_F\_Cluster\_1~4. If you have something more to clarify, please provide detail description.**

|  |  |  |
| --- | --- | --- |
| Company | Answer to issues | comments |
| NTT DOCOMO | Issue Intra\_F\_Cluster\_1: All 3 scenarios are applicable. For all scenarios, the input to the AI/ML model can be from multiple cells (incl. source and neighbouring). The output of the model can also be the prediction of multiple cells. The AI/ML models have advantages in learning the relations among cells compared with legacy methods. Therefore, the cluster-based approach may have more gains.  Issue Intra\_F\_Cluster\_2: Both. Same comment as Question 2.2.2-11.  Issue Intra\_F\_Cluster\_3: IN\_N and OUT\_N can be up to companies report.  Issue Intra\_F\_Cluster\_4: The RSRP differences, which definition is based on our previous suggestion as,  The difference between the L3 cell-level measurement results (measured on the predicted time slots) of the Top-1 predicted cell and the L3 cell-level measurement results of the actual Top-1 cell. |  |
| OPPO | issue 1: It is applicable for all cases but whether we need them all is questionable. For example, if we use L1 measurement to predict L3 results (i.e., Case 3), multiple beam measurements can already provide enough information to determine the postion of UE. Cluster-based approach may not be needed in this case, but would be important for sub case 2 i.e. the input measurement result is L3 cell level measurement  Issue 2: Applicable for co-located cells. In simulation, there is no spatial consistnecy or frequency correlation between non-co-located cells. Using measurements of two non-co-located cells as model input may has the same effect as using one of them as model input, since little additional information will be given to the AI/ML model.  Issue 3: Agree with DCM to leave the number of IN\_N and OUT\_N to companies report. But If we agree the measurements of co-located cells to be model input, then the maximum number of IN\_N and OUT\_N would be no more than 3 for intra-frequency cases and for inter-frequency cases. Plus IN\_N should be >= OUT\_N because model itself will not produce extra information.  Issue 4: yes. For example, for MRRS, at least what is “skipped beams to be measured” and what is “ total beams to be measured” need be further clarified considering the IN\_N and/or OUT\_N could be more than one cell. |  |
| Apple | Can someone (perhaps the proponent?) provide a clear definition of what “cluster” actually means? It’s hard to have this discussion without such definition. |  |
| Mediatek | Issue Intra\_F\_Cluster\_1: Agree with Docomo.  Issue Intra\_F\_Cluster\_2: Since it considers a cluster of cells, it’s can be either co-located or non-co-located.  Issue Intra\_F\_Cluster\_3: IN\_N and OUT\_N can be up to companies report.  Issue Intra\_F\_Cluster\_4: RSRP differences, we can use the RSRP difference between the predict value and the measurement value of the same cell as the starting point. Docomo’s suggestion can also be considered with Top-1 cell prediction.  Our understanding is that the main benefit of considering a cluster-based approach lies in circumventing the need for cell-specific AI/ML models for AI mobility. In a real deployment scenario, the LCM of AI/ML models, which includes tasks such as activation, deactivation, and updates, can become excessively complex and burdensome if each model is tailored to an individual cell. It’s also too early to decide which way e.g., cell-specific or cluster-specific model to go without evaluation on the generalization performance. |  |
| Huawei, HiSilicon |  | We do not think we should study cluster approach. The use cases we have agreed thus far are more than enough to keep us busy in this SID anyway. |
| Samsung | Issue 1: The motivation of using measurement results from multiple cells as model input is to implicitly provide UE’s location information to AI/ML model. Thus, it can be applied to All 3 cases.  Meanwhile, we agree with OPPO’s understanding that it is mainly for the case 2. However, we think the case 2 should be the baseline in RRM measurement prediction for L3 HO enhancement.  Issue 2: It can be applied to both co-located and non-co-located cases. Even though there is no spatial consistency or frequency correlation between non-co-located cells in simulation, the measurement results should be correlated with UE’s location anyway.  Issue 3: Agree with DCM that IN\_N and OUT\_N can be up to companies. However, we are ok to start with the simple cases (i.e., IN\_N = OUT\_N <= 3 and Input SET B cells = Output SET A cells).  Issue 4: Yes.  For the accuracy (i.e., RSRP difference), the dimension of RSRP comparison can be easily expanded to cover the output for multiple cells.  For the measurement reduction metric (i.e., MRRT, MRRS), the definitions above consider only the Intra-cell prediction case. When SET A cells = SET B cells, the metrics can be calculated for each cell in the SET. On the other hand, When SET A SET B, it should be the inter-cell prediction case and we don’t need to evaluate the measurement reduction for the case as discussed in Q. 2.2.2-7. |  |
| vivo | Our initial understanding is that cluster approach can be categorized as inter-cell approach.  Issue 1: all scenarios are applicable  Issue 2: both co-located and non-co-located  Issue 3: the relation between IN\_N and OUT\_N depends on companies’ implementation  Issue 4: the definition of metrics should cover cluster approach |  |
| Ericsson |  | We need to clarify the concept of “cluster” before addressing the issues. |
| Xiaomi | We understand cluster approach means model input includes measurement results from multiple cells. These cells can be a cluster.  With above assumption, the answer is,   1. All scenarios are applicable. 2. Both can be considered. 3. Can be up to companies decision. But IN\_N should be larger than 1. 4. Reuse the RRM measurement KPI. No need to define new metrics. |  |
| CMCC | Agree with Ericsson. |  |
| ZTE | Our understanding of “cluster” is only considered for intrer-cell prediction. We can start from simply case, i.e. to predict one cell first. For the question above:   1. All 3 scenarios are applicable. But we can consider the priority. 2. Both 3. Up to companies, for intra-freq: OUT\_N<= IN\_N; for inter-freq, no restriction. |  |
| Nokia |  | Agree with other companies that more clarification is needed on this cluster-based approach. |
| Intel | For all cluster scenarios, we prefer to keep it with a lower priority. We can restart related discussion once RAN2 made enough progress in the above use cases. |  |

## Simulation assumption

### FR1 and FR2

#### UE trajectory

In section 6.3.1 of TR 38.843, there are three options listed for trajectory modelling:

- Option 1: Linear trajectory model with random direction change.

- UE moving trajectory: UE will move straight 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 moves straight within the time interval with the fixed speed.

- Option 2: Linear trajectory model with random and smooth direction change.

- UE moving trajectory: UE will change the moving direction by multiple steps within an time internal, where the length of the time interval is provided by using an exponential distribution with average interval length, e.g., 5s, with granularity of 100 ms.

- UE moving direction change: At the end of the time interval, UE will change the moving direction with the angle difference A\_diff from the beginning of the time interval, provided by using a uniform distribution within [-45°, 45°].

- The time interval is further broken into N sub-intervals, e.g. 100ms per sub-interval, and at the end of each sub-interval, UE change the direction by the angle of A\_diff/N.

- UE moves straight within the time sub-interval with the fixed speed.

- Option 3: Random direction straight-line trajectories.

- Initial UE location, moving direction and speed: UE is randomly dropped in a cell, and an initial moving direction is randomly selected, with a fixed speed.

- The initial UE location should be randomly drop within the following blue area:



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

- Each sector is a cell and that the cell association is geometry based.

- During the simulation, inter-cell handover or switching should be disabled.

Table 2.3.1-1

Note the UE trajectory in table 2.3.1-1 in RAN1’s simulation is only limited to serving cell. As RAN2 agreed that “Reuse current RAN1’s simulation assumptions as much as possible by extending data generation to neighbouring cells”, this agreement could be also applied for UE trajectory i.e., UE supposes to moves across cells. RAN2 also agreed that “UE trajectory model uses options 1-3 in TR 38.843 section 6.3.1 as the starting point. Down-selection to be discussed in email discussion”. From rapporteur point of view it would desirable that RAN2 can boil down to just one option to easy comparison among companies’ simulation result. Only few contributions show some preference e.g.,[9] propose option1,[10] propose “Prioritize UE trajectory model option 1 of TR 38.843 for low-speed UEs and straight-line trajectory for high-speed UEs”. Another approach could be that RAN2 agree on one option as default one. If company want to use UE trajectory different from default one, it should be provided together with simulation result.

**Question 2.3.1.1-1 How do you think of selection of UE trajectory among 3 options listed in table 2.3.1-1?**

|  |  |  |
| --- | --- | --- |
| Company | Preferred option(s) | comments |
| NTT DOCOMO | Option 3 | Option 3 is preferred to reduce the complexity of the simulation. |
| OPPO | Option 1 | Option 3 makes the prediction easier but is rather simple to model UE moving behavior. Options 1 and 2 can better reflect real UE trajectory. We prefer option 1 because it is simpler than option 2 and achieve a good balance between simulation complexity and realism. |
| Apple | Option 3 or 1 | Slight preference for option 3 (for simplicity), but we can accept option 1.  Furthermore, we think a single option should be agreed. |
| Mediatek |  | No need to down select. We may need to have all the trajectory options to evaluate the generalization performance. |
| Huawei, HiSilicon | Option 1 | Option 3 assumes inter-cell HO does not happen, so should be excluded. Out of the other two options, the results should not differ much between option 1 and 2 while option 1 is simpler. |
| Samsung | Option 3 | Prefer a simple trajectory. |
| vivo | Option 1 | To simulate the actual scenarios, we think that it is necessary for UE to change direction after a relatively long period (e.g., several seconds). Option 3 is too easy to do AI prediction and cannot reflect the performance of AI in actual scenarios. |
| Ericsson | Option 3 | Option 3 is the simplest one and reduce the simulation complexity. |
| Xiaomi | Option 3 as baseline | Option 3 can be baseline. But companies can choose others as well |
| CMCC | Option 1 or Option 2, slightly prefer Option 1 | We prefer the selected option could be applicable for RRM measurement, RLF/HOF and Measurement event predictions. So, UE trajectory could not be limited in one cell. |
| ZTE | Prefer Option 2 | Option 2 is more practical. Since we select Umi for FR2, we should also consider a more challengable UE trajectory for simulation.  Option 3 is also acceptable if the restriction ‘*During the simulation, inter-cell handover or switching should be disabled.*’ is removed. |
| Nokia | 1, 3 | We think O1 can be considered as a starting point. O3 alone is too simple, however we think that this can be modelled together with wrap around and/or bounce back and result in more realistic results. We also think that there is no need to restrict O3 to intra-cell only. |
| Intel | Option 1 as baseline | To reflect more closer to UE trajectory in real-life, we think Option 1 is a good compromise if we want to make sure our simutation evalution work can be useful when considering deployment.  Additionally, for RAN2 mobility simulation, we also needs to consider how UE moves when reach the boundary. |

2nd issue is what should UE do when UE reach boundary of the simulation environment. For example, the simulation environment consists of 2-tier model (7 sites, 3 sectors/cells per site). When a UE is dropped in some way within one of the cell, UE moves with one predefined UE trajectory. When UE reaches the boundary, how should UE do? TR[3] capture two approach in section 5.4.5.1 i.e., wrap-round model and bouncing circle. For the wrap-around model, when the UE hit the simulation border (the wrap-around contour), it will wrap around and enter the simulation area from a different point on the wrap-around contour. For the bouncing-circle model, when the UE hit the simulation border (the bouncing-circle), it will bounce back with a random angle. For the bouncing-circle approach, the simulation area within the bouncing-circle should include 1 tier of complete sites. Only the results from the inner tiers of the sites will be logged, including all the outer border area of the sites.[12] propose 3rd options where the UE trajectory is terminated when UE hits the simulation border. TR[2] also adopt option3. These 3 options are illustrated with a nice Figure in [12].

 

Option 1 Option 2 Option 3

Figure 2.3.1-1: 3 Options for boundary processing [12]

[11] also proposes option 1 and option 2. The main issue for option 1 and 2 is that UE will change UE trajectory suddenly and hence cause sudden change of measurement result of same cell. On the other hand, a UE trajectory with relative long time is also necessary to study mobility performance.

**Question 2.3.1.1-2 How do you think of selection boundary processing as illustrated in Figure 2.3.1-1?**

|  |  |  |
| --- | --- | --- |
| Company | Preferred option(s) | comments |
| NTT DOCOMO | Option 2 or Option 3 | To control the simulation complexity. |
| OPPO | Option 3 | Either option 1 or 2 will have the issue of sudden position or direction change, which will make the AI/ML hard to predict. |
| Apple | Option 3 |  |
| Mediatek | Leave for company implementation | Not sure whether we really need to do this selection. It’s just simulation method. No matter which method to use, the collected data needs to be pre-processed, e.g., remove the data due to UE trajectory sudden change. It can be left to companies to choose. If down-selection is required, we prefer option 2. |
| Huawei, HiSilicon | Option 1 | Probably it is fair to assume that sudden direction changes happen in real life as well. We can of course choose the scenarios where it is easier to predict, but what is the point of doing so if the results cannot be translated into real life scenarios anyway.  Option 3 is acceptable, but the simulation time may be a problem as mentioned by the rapporteur. |
| Samsung | Option 1 | Option 3 is not appropriate at all for mobility performance whose simulation requires longer trajectory e.g. 500m, 1km. Option 3 is only possible for RRM prediction without mobility events.  Option 2 may underestimate interference from other sites, due to the lack of the interference source. It could be ok for RRM prediction, but not suitable for mobility prediction and evaluation. |
| vivo |  | From our view, there is not much performance difference for all the 3 options and we can down-select anyone of them.  And data cleaning for these options is needed to remove the samples that sudden position change occurs during observation/prediction windows. |
| Ericsson | Option 1 | Sudden trajectory change may happen in real world, so this is not an issue |
| Xiaomi | Option 1 |  |
| CMCC | Option 3 | Considering the simulation complexity, we prefer the simple one. |
| ZTE | Option 3 | Option 3 is simple and enough for RRM prediction evaluation. |
| Nokia | 1, 3 | We think its important to consider a trade-off between very simple options (like O3), and more complicated cases, like 1 and 2. This would provide a more realistic view of the gains we expect to see from AIML. We think that at least one of O1 (preferred) or O2 should be considered, since direction changes can happen also in practical deployments. |
| Intel | Option 1 and/or 2 | though simplicity of simiultation work could be considered, but we think UE trajectory in TR 36.839 should be considered as baseline. |

#### Traffic model

In the simulation assumption from contribution [4][5][6][7][8][11], no one select traffic model as simulation parameter. [13] also propose not to consider user plane related performance. Without simulating traffic model, a lot calculation power can be saved during simulation which help make progress in RAN2.

**Question 2.3.1.2-1 Do you agree that no traffic model is simulated in order to evaluate user plane related performance e.g., user throughput?**

|  |  |  |
| --- | --- | --- |
| Company | Position: yes, no | comments |
| NTT DOCOMO | Yes | No user throughput simulations at least for the early stage of this SI. |
| OPPO | Yes | Modelling UE throughput can make the simulation too complicated. R18 AI/ML for air SI saw a few companies doing it. Some metrics, e.g., CDF of RSRP difference, can reflect the impact on user throughput to some extent. |
| Apple | Yes | Throughput simulation can be complicated, other aspects such as scheduler design can make discussion even more complicated. |
| Mediatek | Yes | We can consider it later if time allows. |
| Huawei, HiSilicon | Yes | We see no need to simulate a traffic pattern for the purpose of this study. |
| Samsung | Yes |  |
| vivo | Yes | It’s too complicated to evaluate the throughput. However some simplified KPIs such as the serving cell RSRP or SINR during the simulation time can be considered to show the throughput performance. |
| Ericsson | Yes |  |
| Xiaomi | Yes |  |
| CMCC | Yes | No need to consider traffic model in the evaluation. |
| ZTE | Yes |  |
| Nokia | Yes | We agree with the proposal for intermediate metrics. We can account for this (if needed) in the context of system-level metrics. |
| Intel | Yes |  |

#### UE distribution

There are basically two issues for UE distribution:

Issue1: the possibility for UE to be distributed indoor or outdoor area.

Issue2 how to drop UE into simulation environment.

For issue 1, contribution [6][7][8][12] propose only consider outdoor. If we need consider indoor case, the channel model would also consider indoor scenario or outdoor to indoor scenario. It will make the simulation itself complicated. For 1st study goal, it seems not necessary to have such complexity. For 2nd study goal, there is other parameters e.g. UE speed, or T310 etc. to set up challenging scenario. Maybe it is easy to simply focus on dropping UE outdoor only.

**Question 2.3.1.3-1 Do you agree that UE is dropped 100% outdoor? If no, please clarify your preference**

|  |  |  |
| --- | --- | --- |
| Company | Position: yes, no | comments |
| NTT DOCOMO | Yes (w/ comments) | The baseline case can be 100% outdoor. Suggest not to preclude other options. |
| OPPO | Yes | Agree with DOCOMO. Involvement of indoor scenario may require new AI/ML models that differ from that of 100% outdoor. Indoor scenario can be studied in generalization discussion, if interested. |
| Apple | Yes |  |
| Mediatek | Yes | Agree with Docomo. |
| Huawei, HiSilicon | Yes |  |
| Samsung | Yes | Indoor-outdoor transition will bring another big discussion on channel modeling, and major consideration of mobility is for outdoor. We prefer to focus on the major scenario. |
| vivo | Yes |  |
| Ericsson | Yes | For simplicity. |
| Xiaomi | Yes |  |
| CMCC | Yes | Same view with NTT DOCOMO and OPPO. |
| ZTE | Yes |  |
| Nokia | Yes as baseline | If needed, indoor scenarios can be studied later. |
| Intel | Yes |  |

Contribution [6] also propose 3 options to drop UE:

* Option 1: the UE is randomly dropped within the cell;
* Option 2: the UE is randomly dropped at the edge of cell;
* Option 3: the UE is randomly dropped at the edge of cell and sector;

 

Option 2[6] option 3[6]

Option 1 is adopted in the TR 38.843. By dropping UE at cell and/or sector edge more UEs will experience handover procedure compared to option 1. For RRM measurement prediction use case, it doesn’t matter too much. But it may matter for other use cases related to handover procedure e.g. measurement event prediction etc.

**Question 2.3.1.3-2 How do you think of drop option(s)? If you have other option, please describe the details.**

|  |  |  |
| --- | --- | --- |
| Company | preferred option(s) | comments |
| NTT DOCOMO  (updated) | Option 2 or Option 3  Leave for company implementation | These two options make the simulations more efficient.  Based on the comments, we are also fine with this issue being left to the company's implementation. |
| OPPO | Leave for company implementation | As UE will move a long time in the simulation scenario, its initial position will not have too much impact on performance. |
| Apple | Leave for each company to decide |  |
| Mediatek | Leave for company implementation |  |
| Huawei, HiSilicon | Option 1 | We think that in case simulation time is long enough, we can achieve sufficient number of handovers. We are also OK to leave this up to companies’ preference. |
| Samsung | Option 1 | If we start with RRM prediction, both cell/sector edge and cell center are important. |
| vivo | Option 1 | Option 1 is the simulation assumption we usually use and can meet the needs of all use cases of AI mobility. |
| Ericsson | Option 1, but | It is up to the company to decide which option to use. |
| Xiaomi | Doesn’t matter | Since UE would move, the initial drop position doesn’t make much difference. |
| CMCC | Option 3 | We think it is more in line with the actual situation. |
| ZTE | Option 1 is baseline and Option 3 is optional. | At the early stage of SI, we can take option 1 as baseline. While for system level simulation (to evaluate the handover performance impacts), it’s up to company to decide whether to use option 3 to reduce the simulation time. |
| Nokia | Option 1 | We are also ok to leave this for each company to choose. |
| Intel | Option 1 |  |

#### UE speed

As for UE speed, the candidate value is 3,30,60,90,120 Km/h. For 1st study goal, it doesn’t make too much sense to evaluate high speed. While for 2nd study goal, comparison between different speed is helpful to understand the performance gain brough by model in different speed. Since UE speed is a critical parameter for simulation on mobility performance, it would be desirable that company have common understanding which UE speed(s) should be evaluated.

**Question 2.3.1.4-1 Which UE speeds among 3,30,60,90,120 Km/h are chosen for which handover scenario (FR1\_to\_FR1, FR2\_to\_FR2) and for what purpose (e.g., study goal1, study goal 2)? Note selected UE speeds could be sub set or full set of the listed ones.**

|  |  |  |
| --- | --- | --- |
| Company | chosen set of UE speeds and corresponding handover scenario, purpose | comments |
| NTT DOCOMO | 30, 60, 90, 120 kmph as common assumptions. | 3 kmph may be used only if the indoor UEs are considered. |
| OPPO | For 1st study goal: 30 km/h as baseline and open for 3 and 60 .  For 2nd study goal: 120km/h as baseline, and open for 60 and 90 km/h. | For 1st goal, we don’t hink high speed like 90 and 120Km/h are necessary since we expect the prediction accuracy is not promising  For 2nd goal, we don’t think low speed like 3,30km/h is challenging enough to show the potential of AIML |
| Apple | Agree with OPPO |  |
| Mediatek | 30, 60, 90, 120 kmph as common assumptions. | We need different speed setting for generalization performance evaluation. |
| Huawei, HiSilicon | For study goal 1: 30, 60, 120  For study goal 2: 120 | We need to limit somehow the simulated cases. The lowest speed is not interesting for neither scenario while for HO performance improvement, we should focus on high mobility scenario where the KPI improvement is most needed.  We wouldn’t like to exclude high speeds for FR1 just based on “the feeling”. The purpose of the study is to evaluate the accuracy in different conditions, we shouldn’t jump to conclusions already. |
| Samsung | Start with 30km/h, can add 60km/h later | Prefer to start with a common major scenario. |
| vivo | 60, 90, 120 km/h for study goal 2, and 30, 60,90, 120 km/h for study goal 1. | For study goal 2, the baseline handover failure rate of 30km/h is relatively low from our simulation results [12]. Therefore, we prefer to focus on high speeds such as 60, 90 and 120 km/h. |
| Ericsson | 3, 30, 60, 90, 120 Kmph are ok to consider | We should not exclude any speed at this stage of the study. |
| Xiaomi | 3, 30, 120 km/h | 3 for pedestrian, 30 for low speed vehicle, 120 for high speed vehicle. |
| CMCC | 30km/h, 60km/h, 90km/h, 120km/h as common assumptions |  |
| ZTE | Agree with OPPO. |  |
| Nokia | 3, 30, 60, 90, 120 Kmph | We think it’s too early to downselect among the different speed patterns as we will need to consider them when discussing generalization aspects of the models anyway. Nevertheless, if there is a need to downselect, we think that for goal 1, 30 kmph would be more suitable as baseline, whereas for goal 2, 120 kmph could be considered. |
| Intel | 30 km/h as baseline, other speed, e.g. 60, 90, 120 kmph are optional. |  |

#### Channel modelling

Contribution [14] raised few issues w.r.t. channel model additionally. [14] has propose 8 “For simplified simulation, fast-fading model is optional, whether to adopt it is up to each company.” The modelling of fast fading can be time-consuming and has limited impact on cell-level results that have been L1/L3 filtered. On the other hand, L1 beam level measurement is input parameter in RRM sub case 1 and 3. And so far, it is not clear whether it should be L1 filtered or not. In addition, for RLF/HOF evaluation, L1 raw data (before L1 filtering) with fast fading is expected to reflect the variation of wireless channel. The concern on work load is at the phase to generate dataset instead of running simulation phase. If fast fading is deemed necessary for RLF/HOF evaluation, then it could be also used for RRM measurement prediction since anyway it will be there. Another approach is that for RRM measurement prediction use case, fast fading may be optional . If it is necessary for RLF/HOF evaluation, it can be added on top of agreed simulation assumption.

**Question 2.3.1.5-1 In which use case(s)/sub-use case(s), do you think that fast-fading model is necessary?**

|  |  |  |
| --- | --- | --- |
| Company | Position: yes or no | comments |
| NTT DOCOMO | Yes | At least for RRM measurement prediction, since the channel will be too deterministic without the fast-fading model. |
| OPPO | Yes with comment. | It mainly depends on the type of input measurement. For sub case 1 and 3, it make sense,fast-fading is needed to reflect the fluctuation of wireless channels. For sub case 2, it doesn’t matter because fluctuation could be smoothed by L3 filtering . |
| Apple | No | We are indeed concerned we will be spending more time in this study on simulations rather than AI/ML, which this study is supposed to be about.  Furthermore, [2] wasn’t even discussed in RAN2#125bis – if we are to have this discussion, we should start it online. |
| Mediatek | Yes | We need to consider the fast fading model for all use cases. |
| Huawei, HiSilicon | Yes | We think fast fading is an important part of channel model and it should be part of simulation work. Otherwise we may get good results but which will never translate into real life scenarios. |
| Samsung | Not essential for cell-level mobility | For case 1/3 beam prediction, fast fading may be needed. But for cell-level mobility, fast fading has been considered as random noise which should be eliminated by L1/L3 filtering. We see that it just increases the simulation complexity. |
| vivo | Yes | We should try our best to simulate the actual environment to ensure that AI methods we study can really bring gains to real-life networks. |
| Ericsson | Yes, for all 3 sub-cases. |  |
| Xiaomi | Not necessary | Agree with Samsung. |
| CMCC | Yes | It is necessary, especially for the 2nd study goal. |
| ZTE | Yes | Fast fading model is needed for all the use cases, otherwise, it cannot reflect the real enviroment and the performance impact caused by AI. |
| Nokia | Yes for all use/sub-use cases | We think that fast fading components would be present in practical deployments. Considering it in our evaluations for all the use-cases is needed to ensure that the results are realistic. |
| Intel |  | We prefer to start from simple assumption for channel model. |

[14] also propose following models in TR38.901[15] are not considered to simplify the channel modelling for RAN2:

- Oxygen absorption (7.6.1 of TR 38.901)

- Large bandwidth and large antenna array (7.6.2)

- Time-varying Doppler shift (7.6.6)

- UT rotation (7.6.7)

- Explicit ground reflection model (7.6.8)

- Blockage (7.6.4)

RAN2 already agreed not to consider UE(UT) rotation. In addition, bandwidth and antenna array are covered by system bandwidth and gNB/UE antenna configuration parameters and hence not discussed over here.

**Question 2.3.1.5-2 Do you agree to not consider Oxygen absorption (7.6.1), Time-varying Doppler shift (7.6.6), Explicit ground reflection model (7.6.8) and blockage (7.6.4)? If you have any further model to be skipped by RAN2, please provide it with detail comments.**

|  |  |  |
| --- | --- | --- |
| Company | Position: yes or no | comments |
| NTT DOCOMO | No | Firstly, the Time-varying Doppler shift is necessary since we are simulating UE mobility across a long distance, the Doppler shift changes with the UE position.  Second, the blockage can be optionally adopted, which is a practical case for causing HO in real-life networks. |
| OPPO | Yes | They are optional functions in 38.901. We can simply drop these additional components during initial simulation to make it easier.  Time-varying doppler shift is an addition component of small-scale factor that can be averaged in frequency-domain when getting RSRP. Therefore, it is not a must. |
| Apple | Yes | In Rel-18 AI/ML study, such modeling aspects were not considered by all companies. At most, it is optional for a company to report the inclusion of any modeling aspects, but the baseline is without any of them. |
| Mediatek | Yes | We can take these assumptions as starting point for simulation. |
| Huawei, HiSilicon | Yes |  |
| Samsung | Yes | Even in most of RAN1 simulations, those are not used. |
| vivo | Yes with comments | We agree with NTT DOCOMO that blockage can be optionally adopted. |
| Ericsson | Yes | For simplicity. |
| Xiaomi | Yes | We don’t need to consider these aspects. |
| CMCC | Yes | We prefer to simplify the channel modelling in RAN2. |
| ZTE | Yes |  |
| Nokia | Yes | Agree that this would simplify the simulation effort. |
| Intel | Yes with comment | Agree with Apple that it should be optional and baseline can be without any of them. |

One more issue raised by from [14] is spatial dependency during LOS and NLOS transition. [14] believe that FR2 band is easily broken, and means that mobility performance highly depends on the LOS-NLOS transition. And TR 38.901 defines modelling methodology of LOS-NLOS transition called LOSsoft, as a spatial consistency model. LOSsoft state is an intermediate state between transitions which depend on correlation distance and frequency band. And [14] propose “RAN2 shall consider spatial consistency of LOS-NLOS transition, according to TR 38.901”. Note it is also complicated to add this model into channel modelling. So, there is trade-off between complexity and performance gain. Since it is so far proposed only by one company, there could be 3 options:

Option 1: it is mandatory in the channel modelling ([14]’s proposal)

Option 2: it is optional in the channel modelling

Option 3: it is not considered in the channel modelling

**Question 2.3.1.5-3 Which option do you prefer in terms of LOSsoft?**

|  |  |  |
| --- | --- | --- |
| Company | Position: option1, option2, option3 | comments |
| NTT DOCOMO | Option 2 | It should be encouraged to model it. |
| OPPO | Option 2 | In 38.901, it states that “To circumvent such hard transitions the optional soft LOS state can be considered...”. It is clear soft LOS is an optional function in channel modelling and can be left for company implementation. |
| Apple | Option 2 |  |
| Mediatek | Option 1, option 2 | While LOFsoft is an optional feature in channel modelling, its inclusion and subsequent performance evaluation are beneficial for gaining a comprehensive understanding of the enhancements that AI can achieve. |
| Huawei, HiSilicon | Option 3 | We think it is sufficient to rely on the LOS probability formulation in TR 38.901. |
| Samsung | Option 1  (at least highly encourage to model it) | The reason why TR 38.901 says it’s optional is that traditional RAN1 system-level simulation does not consider UE mobility at all, except for Doppler shift. Also, distance-dependent LOS probability changes with BS-UT distance. If we do not consider this, UE’s LOS state is either always LOS or always NLOS. It does not make sense.  Another consideration is that sudden RSRP change in FR2 occur due to LOS -> NLOS transition. It is a major challenge in FR2 mobility. If we do not consider this aspect, it will not be a proper FR2 modeling at all. |
| vivo | Option 2 | Soft los feature is useful for mobility evaluation, but it may double the simulation time (e.g., both the los channel and nlos channel needs to be calculated to get the channel of soft los for each time of channel update according to TR 38.901). |
| Ericsson | Option 2 |  |
| Xiaomi | Option 3 |  |
| CMCC | Option 2 | It is up to company implementation. |
| ZTE | Option 2 |  |
| Nokia | Option 2 | We do not see the need to make it mandatory, however are ok to not preclude it entirely from the study as predicting LOS/NLOS transitions is a key challenge in radio measurement-based ML models. Companies can bring evaluations for NLOS if they see significant gains. |
| Intel | Option 2 | it should be optional, companies are welcome to bring evaluation results on this aspects. |

### FR2

Contributions [4][5][6][7][8][11] list detail simulation assumptions. The cross check among those contributions shows that some of the parameters are not necessary for RAN2 simulation. Table 2.3.4-1 list the parameters which are chosen by all or majority of the previous contributions from table 6.3.1-1[2].

|  |  |
| --- | --- |
| Parameter | Value |
| Frequency Range | FR2 @ 30 GHz; SCS: 120 kHz |
| Deployment | 200m ISD, 2-tier model with wrap-around (7 sites, 3 sectors/cells per site)  Other deployment assumption is not precluded |
| Channel model | UMa with distance-dependent LoS probability function defined in Table 7.4.2-1 in TR 38.901. |
| System BW | 80MHz |
| UE Speed | For spatial domain beam prediction: 3km/h  For time domain beam prediction: 30km/h (baseline), 60km/h (optional) 90km/h (optional), 120km/h (optional)  Other values are not precluded |
| UE distribution | 10 UEs per sector/cell for system performance related KPI (if supported) [e.g., throughput] for full buffer traffic (if supported) evaluation (model inference).  X UEs per sector/cell for system performance related KPI for FTP traffic (if supported) evaluation (model inference).  Other values are not precluded.  Number of UEs per sector/cell during data collection (training/testing) is reported by companies if relevant.  For spatial domain beam prediction (optional to compare different UE distributions assumptions):  - Option 1: 80% indoor ,20% outdoor as in TR 38.901  - Option 2: 100% outdoor  For time domain prediction: 100% outdoor |
| BS Antenna Configuration | Antenna setup and port layouts at gNB: (4, 8, 2, 1, 1, 1, 1), (dV, dH) = (0.5, 0.5) λ  Other assumptions are not precluded.    Companies to explain TXRU weights mapping.  Companies to explain beam selection.  Number of BS beams: 32 or 64 downlink Tx beams (max number of available beams) at NW side. Other values, e.g., 256 not precluded. |
| BS Antenna radiation pattern | TR 38.802 Table A.2.1-6, Table A.2.1-7 |
| UE Antenna Configuration | Antenna setup and port layouts at UE: (1, 4, 2, 1, 2, 1, 1), 2 panels (left, right)  Other assumptions are not precluded  Companies to explain TXRU weights mapping.  Companies to explain beam and panel selection.  Number of UE beams: 4 or 8 downlink Rx beams (max number of available beams) per UE panel at UE side. Other values, e.g., 16 not precluded. |
| UE Antenna radiation pattern | TR 38.802 Table A.2.1-8, Table A.2.1-10 |
| BS Tx Power | 40 dBm (baseline)  Other values (e.g., 34 dBm) not precluded |
| Maximum UE Tx Power | 23 dBm |
| BS receiver Noise Figure | 7 dB |
| UE receiver Noise Figure | 10 dB |
| Inter site distance | 200 m |
| BS Antenna height | 25 m |
| UE Antenna height | 1.5 m |
| Spatial consistency | At least for BM-Case1, companies report the one of spatial consistency procedures:  - Procedure A in TR38.901  - Procedure B in TR38.901 |
| UE trajectory model | Please check section 2.3.1 |

Table 2.3.4-1

Since no one mention whether any parameter could be different between UE and network sided model, as starting point, parameters in the table 2.3.3-1 is assumed common for both UE sided model and network sided model unless otherwise described by rapporteur.

**Question 2.3.4-1 Do you agree to take simulation parameter in table 2.3.4-1 as starting point for both UE sided model and network sided model? If you have different opinion, please provide your detail comments.**

*Note detail value will be discussed in later questions i.e., here the focus is to remove or to add parameters and whether parameter is UE or network sided model specific.*

|  |  |  |
| --- | --- | --- |
| Company | Position: yes or no | comments |
| NTT DOCOMO | Yes |  |
| OPPO | Yes |  |
| Apple | Yes |  |
| Mediatek | Yes |  |
| Huawei, HiSilicon | Yes to most, but see comments | For channel bandwidth we think 100 MHz is a more reasonable value considering real life deployments. |
| Samsung | Yes |  |
| vivo | Yes |  |
| Ericsson | Yes |  |
| Xiaomi | Yes |  |
| CMCC | Yes |  |
| ZTE | Yes to most, see comments | For BS Antenna radiation pattern, remove Table A.2.1-7, because table A.2.1-7 is for Indoor BS antenna radiation pattern for above 6GHz;  For UE Antenna radiation pattern, remove Table A.2.1-10, because table A.2.1-10 is for HST scenario. |
| Nokia | Yes to most, but see comment | We think the number of beams in the GoB can be left to the companies to select (considering the increased simulation runtime associated with very high number of beams). |
| Intel | Yes | ok with 100MHz as proposed by HW for real-life deployment |

There are proposals from company about detail values. We will discuss them one by one. W.r.t. frequency range, RAN2 agreed that “For FR2, only FR2-1 is considered, e.g., band n257. 30GHz central frequency can be adopted to reuse RAN1’s work as much as possible. FFS any other band”. Only contribution [8] propose 28GHz for FR2-1. Rapporteur believe it is not necessary to add one more frequency just due to such minority view.

**Question 2.3.4-2 Do you agree for FR2-1, only 30GHz is adopted as central frequency?**

|  |  |  |
| --- | --- | --- |
| Company | Position: yes or no | comments |
| NTT DOCOMO | Yes |  |
| OPPO | Yes |  |
| Apple | Yes |  |
| Huawei, HiSilicon | Yes |  |
| Samsung | Yes | No strong view |
| vivo | Yes |  |
| Ericsson | Yes |  |
| Xiaomi | Yes |  |
| CMCC | Yes |  |
| ZTE | Yes |  |
| Nokia | Yes |  |
| Intel | Yes |  |

For inter-site distance, majority company e.g.,[6][7][8] prefer 200m for FR2-1 apart from [4], which propose 500ms. FR2 is potential frequency range to be evaluated for 2nd study goal i.e., to improve handover performance. ISD with 500m can’t set up a challenging scenario for such study goal.

**Question 2.3.4-3 Do you agree that ISD for FR2 should be 200m? If you have different opinion, please provide detail value.**

|  |  |  |
| --- | --- | --- |
| Company | Position: yes or no | comments |
| NTT DOCOMO | Yes |  |
| OPPO | Yes |  |
| Apple | Yes |  |
| Mediatek | Yes |  |
| Huawei, HiSilicon | Yes |  |
| Samsung | Yes | We are fine to have a common ISD for all sceanrios |
| vivo | Yes |  |
| Ericsson | Yes |  |
| Xiaomi | Yes |  |
| CMCC | Yes |  |
| ZTE | Yes |  |
| Nokia | Yes |  |
| Intel | Yes |  |

As for channel modelling, RAN2 agreed “focus on Urban Macro (UMa) for FR1 and Umi for FR2”. By combining this agreement with value of “channel model” in table 2.3.4-1, the starting point for FR2 could be “UMi with distance-dependent LoS probability function defined in Table 7.4.2-1 in TR 38.901”.

**Question 2.3.4-4 Do you agree that the baseline channel model for FR2 is defined as “UMi with distance-dependent LoS probability function defined in Table 7.4.2-1 in TR 38.901”?**

|  |  |  |
| --- | --- | --- |
| Company | Position: yes or no | comments |
| NTT DOCOMO | Yes |  |
| OPPO | Yes |  |
| Apple | Yes |  |
| Mediatek | Yes |  |
| Huawei, HiSilicon | Yes |  |
| Samsung | Yes | We see FR2 with more signal fluctuation is the most challenging scenario of mobility performance. UMi is for such scenario. |
| vivo | Yes |  |
| Ericsson | Yes |  |
| Xiaomi | Yes |  |
| CMCC | Yes |  |
| ZTE | Yes |  |
| Nokia | Yes |  |
| Intel | Yes |  |

Apart from uncertain parameters in table 2.3.4-1. There may be some other parameters to be discussed by companies.

**Question 2.3.4-5 Do you have any other parameters to be discussed? If so, please provide detail description and reason behind.**

|  |  |  |
| --- | --- | --- |
| Company | Any other parameters? | comments |
| vivo | Handover and RLM parameters. | Handover and RLM related parameters need to be discussed if we reach the consensus that system-level performance should also be considered for RRM prediction. |

### FR1

At last RAN2 meeting, it was agreed that both FR1 and FR2 should be evaluated. As for simulation assumption there is no parameters was agreed as starting point for FR1 apart from propagation scenario (Uma i.e., Urban macro cell) and one central frequency (4GHz). It could be difficult to draft the simulation parameters from sketch. Rapporteur noticed that table 6.2.1-1 [2] is simulation assumptions for evaluation of CSI feedback, where the propagation scenario is Uma and the central frequency could be 4GHz. Some company also provide detail simulation assumptions for FR1. We can have a baseline table based on table 6.2.1-1 and company’s contribution.

**Question 2.3.3-1 Do you agree table 6.2.1-1 is taken as starting point for FR1 simulation assumptions?**

|  |  |  |
| --- | --- | --- |
| Company | Position: yes or no | comments |
| NTT DOCOMO | Yes(w/ comments) | It is fine to use it as a template, but many revisions are necessary, and some items in the table are not necessary for our studies.  We also suggest adding a parameter beam number to the table, which candidate value can be 4. |
| OPPO | Yes | Revisions of the table will be discussed in the following. |
| Apple | Yes |  |
| Mediate | Yes | Revison should be allowed. |
| Huawei, HiSilicon | Yes |  |
| Samsung | Yes |  |
| vivo |  | Simulation assumption in Table A.2.5-2 of TR 38.802 is used for evaluation of beam management and may be more appropriate for mobility evaluation than using the CSI feedback simulation assumption. |
| Ericsson | Yes |  |
| CMCC | Yes |  |
| ZTE | Yes | Some revisions are needed since the scenario in table 6.2.1-1 is dense urban. |
| Nokia | Yes | Ok as starting point. Potential revisions should not be precluded. |
| Intel | Yes |  |

Then we start to discuss important assumptions. For FR1, RAN2 agreed that “For FR1, band n77/n78 is considered with 4GHz as the central frequency. FFS any other band”. The FFS is mainly for another frequency for inter-frequency scenario. 2GHz is proposed by [6] and RAN1 usually also use 2GHz as FR1 frequency which is also reflected in table 6.2.1-1[2]. As for SCS, [5][8] propose to use 30KHz while [6] propose 15KHz. In real deployment 2GHz is more likely configured with 15KHz while 4GHz is more likely configured with 30KHz.

**Question 2.3.3-2 Do you agree to take {4GHz,30KHz} as frequency for intra-frequency scenario and {2GHz, 15KHz} as another frequency for inter-frequency scenario?**

|  |  |  |
| --- | --- | --- |
| Company | Position: yes or no | comments |
| NTT DOCOMO | No | For the inter-frequency scenario, we suggest considering two options. One could be 2GHz, as suggested, which is far from Band 1 (4GHz). Another could be a frequency not so far from the Band 1, such as 3.5GHz. With these two options, we can check the width of the frequency gap so that AI/ML can make a good prediction. |
| OPPO | Yes | 4GHz has already been agreed in RAN2#125bis and 2GHz is also commonly adopted by FR1 simulations in TR 38.843 and TR 36.839 |
| Apple | Yes |  |
| Mediatek |  | We require additional information from network vendors and operators regarding FR1 deployment. We are open to considering more realistic deployment scenarios based on their insights. |
| Huawei, HiSilicon | Yes |  |
| Samsung | Yes |  |
| vivo | Yes |  |
| Ericsson | Yes |  |
| CMCC | Yes with comments | If table 6.2.1-1 [2] is taken as the baseline for FR1 simulation assumptions, both 2GHz and 4GHz should be considered for intra-frequency scenario. |
| ZTE | Yes |  |
| Nokia | Yes, but see comments | Agree with the central frequency values. Regarding the SCS we think the proposed values can be used as baseline, however other combination of values should also not be precluded as they would likely be needed when doing the generalization analysis of the models. |
| Intel |  | we also suggest to wait for operators’ input on the other center frequency for inter-frequency scenario. |

The deployment of FR1 could be same as FR2 i.e.,“2-tier model (7 sites, 3 sectors/cells per site)”. Thus could help to reduce simulation work load.

**Question 2.3.3-3 Do you agree that FR1 take the same deployment as FR2 i.e. to set up 2-tier model (7 sites, 3 sectors/cells per site)?**

|  |  |  |
| --- | --- | --- |
| Company | Position: yes or no | comments |
| NTT DOCOMO | Yes |  |
| OPPO | Yes | Adopting the same deployment model can reduce companies’ workload. |
| Apple | Yes |  |
| Mediatek | Yes |  |
| Huawei, HiSilicon | Yes |  |
| Samsung | Yes |  |
| vivo | Yes |  |
| Ericsson | Yes |  |
| Xiaomi | Yes |  |
| CMCC | Yes |  |
| ZTE | Yes |  |
| Nokia | Yes |  |
| Intel | Yes |  |

The ISD in current table 6.2.1-1[2] is 200m. Based on contributions from company, it is likely that ISD of FR2 is 200m. Considering FR1 is usually for coverage purpose and the evaluation of FR1 is targeting 1st study goal, the ISD for FR1 could be more relaxed compared to FR2. In addition, contribution [5][6] propose ISD of FR1 is 500m.

**Question 2.3.3-3a Do you agree that ISD of FR1 is 500m? If no, please provide suggested value**

|  |  |  |
| --- | --- | --- |
| Company | Position: yes or no | comments |
| NTT DOCOMO | No | 200m is more typical and widely used. |
| OPPO | Yes |  |
| Apple | Yes |  |
| Mediatek |  | We require additional information from network vendors and operators regarding FR1 deployment. We are open to considering more realistic deployment scenarios based on their insights. |
| Huawei, HiSilicon | Yes |  |
| Samsung | No | We prefer to have a common ISD. |
| vivo | Yes |  |
| Ericsson | Yes |  |
| CMCC | Yes |  |
| ZTE | Yes | According to Table 6.1.4-1 in TR 38.913, 500ms is the typical ISD for UMa. |
| Nokia | No | Agree with Docomo, Samsung and prefer to keep the usual value of 200m. Larger ISD values would make simulations longer (to obtain sufficient number of HO), especially when evaluating failure prediction use-cases in FR1. We are also ok to leave this parameter to be selected by each company |
| Intel |  | 200m ISD is considered as baseline, other ISD values waiting for input from operators. |

As for the channel modelling, RAN2 agreed that “focus on Urban Macro (UMa) for FR1 and Umi for FR2”. So the recommended channel modelling is “UMi with distance-dependent LoS probability function defined in Table 7.4.2-1 in TR 38.901”

**Question 2.3.3-4 Do you agree that channel modelling of FR1 is “UMa with distance-dependent LoS probability function defined in Table 7.4.2-1 in TR 38.901”?**

|  |  |  |
| --- | --- | --- |
| Company | Position: yes or no | comments |
| NTT DOCOMO | Yes |  |
| Apple | Yes |  |
| Mediatek | Yes |  |
| Huawei, HiSilicon | Yes |  |
| Samsung | Yes | LOS probability highly depends on BS-UT distance. It should be considered mainly for interference modelling and inter-cell modelling. |
| vivo | Yes |  |
| Ericsson | Yes |  |
| Xiaomi | Yes |  |
| CMCC | Yes |  |
| ZTE | Yes |  |
| Nokia | Yes |  |
| Intel | Yes |  |

There are both 10 and 20MHz in current table 6.2.1-1[2]. 20MHz is proposed by [5][6]. Rapporteur think one bandwidth should be sufficient for evaluation.

**Question 2.3.3-5 Do you agree that system bandwidth for FR1 is 20MHz? If no, please provide your suggested bandwidth**

|  |  |  |
| --- | --- | --- |
| Company | Position: yes or no | comments |
| NTT DOCOMO | Yes |  |
| OPPO | Yes |  |
| Apple | Yes |  |
| Mediatek | Yes | 10Mhz can also be considered. |
| Huawei, HiSilicon | Yes |  |
| Samsung | Yes |  |
| vivo | Yes |  |
| Ericsson | Yes |  |
| Xiaomi | Yes |  |
| CMCC | Yes |  |
| ZTE | Yes |  |
| Nokia | Yes |  |
| Intel | Yes |  |

For parameters which is missed from table 6.2.1-1[2], a value is recommended by rapporteur based on contributions at last RAN2 meeting.

|  |  |  |
| --- | --- | --- |
| Parameter | Value | comment |
| Frequency Range | FR1 only, 2GHz as baseline, optional for 4GHz (if R16 as baseline)  FR1 only, 2GHz with duplexing gap of 200MHz between DL and UL, optional for 4GHz (if R17 as baseline) | Up to Question 2.3.3-2 |
| Deployment | Dense Urban (Macro only) is a baseline.  Other scenarios (e.g., UMi@4GHz 2GHz, Urban Macro) are not precluded. | Up to Question 2.3.3-3 |
| Channel model | According to TR 38.901 | Up to Question 2.3.3-4 |
| System BW | 10 MHz for 15kHz as a baseline, and configurations which emulate larger BW, e.g., same sub-band size as 40/100 MHz with 30kHz, may be optionally considered. Above 15kHz is replaced with 30kHz SCS for 4GHz (if R16 as baseline)  20 MHz for 15kHz as a baseline (optional for 10 MHz with 15KHz), and configurations which emulate larger BW, e.g., same sub-band size as 40/100 MHz with 30kHz, may be optionally considered. Above 15kHz is replaced with 30kHz SCS for 4GHz (if R17 as baseline) | Up to Question 2.3.3-5 |
| UE Speed |  | Up to question 2.3.1.4 |
| UE distribution | CSI compression: 80% indoor (3 km/h), 20% outdoor (30 km/h)  CSI prediction: 100% outdoor (10, 20, 30, 60, 120 km/h) including outdoor-to-indoor car penetration loss per TR 38.901 if the simulation assumes UEs inside vehicles. No explicit trajectory modeling considered for evaluations.please check question2.3.1.3-1 | Up to question 2.3.1.3 |
| BS Antenna Configuration | Companies need to report which option(s) are used between  - 32 ports: (8,8,2,1,1,2,8), (dH,dV) = (0.5, 0.8)λ  - 16 ports: (8,4,2,1,1,2,4), (dH,dV) = (0.5, 0.8)λ  Other configurations are not precluded. | No change |
| BS Antenna radiation pattern | 3-sector antenna radiation pattern, 8 dBi | Proposed by [5] |
| UE Antenna Configuration | 4RX: (1,2,2,1,1,1,2), (dH,dV) = (0.5, 0.5)λ for (rank 1-4)  2RX: (1,1,2,1,1,1,1), (dH,dV) = (0.5, 0.5)λ for (rank 1,2)  Other configuration is not precluded. |  |
| UE Antenna radiation pattern | Omni-direction | Proposed by [5][8] |
| BS Tx Power | 41 dBm for 10MHz, 44dBm for 20MHz, 47dBm for 40MHz | Up to Question 2.3.3-5 |
| Maximum UE Tx Power | 23dbm | Proposed by [4][5] |
| BS receiver Noise Figure | 5db | Proposed by [5][6] |
| UE receiver Noise Figure | 9dB |  |
| Inter site distance | 200m | Up to Question 2.3.3-3 |
| BS Antenna height | 25m |  |
| UE Antenna height | Follow TR36.873, which is 1.5m | Proposed by [5] |
| Spatial consistency | At least for BM-Case1, companies report the one of spatial consistency procedures:  - Procedure A in TR38.901  - Procedure B in TR38.901 | Same as FR2, which recommended by rapporteur |
| UE trajectory model |  | Up to Question 2.3.1.1 |

Table 2.3.3-1

**Question 2.3.3-6 Do you agree the recommended value for parameters with yellow colour in table 2.3.3-1 for FR1? If you have different opinion, please provide your comment**

|  |  |  |
| --- | --- | --- |
| Company | Position: yes or no | comments |
| NTT DOCOMO | Yes |  |
| OPPO | Yes |  |
| Apple | Yes |  |
| Mediatek | Yes |  |
| Huawei, HiSilicon | Yes |  |
| Samsung | Yes |  |
| vivo | Yes |  |
| Ericsson | Yes |  |
| Xiaomi | Yes |  |
| CMCC | Yes |  |
| ZTE | Yes | Some words need to be updated.  For spatial consistency, the word ‘BM-case 1’ is not suitable for AI mobility, and can be replaced by spatial domain measurement prediction. |
| Nokia | Yes |  |
| Intel | Yes |  |

#### FR1 inter-frequency specific

About inter-frequency correlation,[14] propose to consider section 7.6.5 of TR[15]. Contribution [8] propose few detail proposals. Rapporteur’s understand is that those proposals are aligned with basic principle in section 7.6.5[15].

**Question 2.3.3.1-1: Do you agree section 7.6.5 [15] is taken as baseline for inter-frequency correlation model?**

|  |  |  |
| --- | --- | --- |
| Company | Position: yes or no | comments |
| NTT DOCOMO | Yes |  |
| OPPO | Yes |  |
| Apple | Yes |  |
| Mediatek | Yes |  |
| Huawei, HiSilicon | No | We think that the baseline can be to have no correlation assumptions and optionally companies can follow the guidelines in 7.6.5 of [15]. |
| Samsung | Yes |  |
| vivo | Yes |  |
| Ericsson | Yes |  |
| Xiaomi | Yes |  |
| CMCC | Yes |  |
| ZTE | Yes, but some update is needed. | At least the following update needs to be considered:  Based on TR 38.901, cluster specific shadowing fading is modeled as random function independent of frequency. While, in the reality, the adjacent frequency may suffer the similar shadowing fading, so in out understanding, random function model independent of frequency may not be suitable. |
| Nokia | No | Agree with HW. |

### RRC parameters

To train model for RRM prediction use case, genie L3 cell level measurement result should be generated as label. In addition, for RRM sub case 1, the predicted L1 beam level measurement need be postprocessed so that a predicted L3 cell level measurement can be produced. That’s why RRC parameters related to consolidation and L3 filtering should be aligned among company. Contribution [7] also propose to align measurement gap configuration for inter-frequency scenario.

**Question 2.3.4-1: Do you agree to setup following RRC parameters as simulation assumption? If you have other parameters to recommend, please provide detail description.**

* RRC parameters for measurement consolidation
* RRC parameters for L3 filtering
* Measurement gap configuration

|  |  |  |
| --- | --- | --- |
| Company | Position: yes or no | comments |
| NTT DOCOMO | Yes(w/ comments) | The aligned parameters can be used for the baseline schemes. For the study on the overhead reduction, these parameters may be adjusted. |
| OPPO | Yes |  |
| Apple | Yes |  |
| Mediatek | Yes | For system level simulation, we also need to consider RLM related parameters. |
| Huawei, HiSilicon | Yes |  |
| Samsung | Yes |  |
| vivo | Yes with comments | We think the L1 filtering parameter (e.g., the number of measurement samples used for averaging) should also be aligned for the simulation, although it does not belong to RRC parameter. |
| Ericsson | Yes |  |
| Xiaomi | Comments on gap | We understand the gap should be equal to sampling period. Since sampling period has been aligned with companies according to **Question 2.2.2-5**, gap configuration is also aligned. |
| CMCC | Yes |  |
| ZTE | See comments | Yes for measurement consolidation RRC parameters;  No for L3 filtering RRC parameters, since we think L3 filtering is not needed, or to only consider Ki =0 in the RRM measurement prediction. See our reply to Question 2.2.1-1.  For gap configuration, it needs further discussion and clarification. |
| Nokia | Yes |  |
| Intel | Yes |  |

### Applicability of simulation assumption

So far, the simulation assumptions discussion is based on the RRM measurement prediction use case. However RAN2 is planning discuss other use cases in future meeting including RAN2#126. If we can identify the common simulation assumption for all use cases as much as possible then it would save time to re-open the discussion again. And in future meeting RAN2 can focus on delta part i.e. something to add on or some parameter to be updated.

**Question 2.3.5-1: Which simulation assumptions discussed in section 2.3.1~2.3.3 are common for all use cases? If you identify that an assumption is use case specific, please explain the details.**

|  |  |  |
| --- | --- | --- |
| Company | Common simulation assumptions | Use case specific assumptions |
| OPPO | all of them |  |
| Apple | Too early to decide |  |
| Huawei, HiSilicon | We think we can reuse those assumptions for all use cases as a baseline, but for some use cases we should limit the evaluated scenarios, e.g. there is no need to evaluate RLF/HOF for FR1 and low speeds. |  |
| Samsung | Most of them, depending on the conclusion of this discussion | For event/HOF/RLF predictions, we should consider longer trajectory and prediction window. Also, LOSsoft and wrap-around processing should be considered. |
| Ericsson | The simulation assumptions discussed previously can be used as baseline. However some changes might be needed (too early to decide now). |  |
| Xiaomi | Try to reuse as much as possible. But agree it may be too early to conclude without RLF/HOF modelling. |  |
| CMCC | Same view with Ericsson. |  |
| ZTE | Same view as Ericsson. |  |
| Nokia | Agree with other companies that its too early to decide |  |
| Intel | all of them with some limitation on scenario to certain use cases |  |

# Conclusion

# Reference

1. RP-234055 Study on Artificial Intelligence (AI)/Machine Learning (ML) for mobility in NR
2. TR 38.843 Study on Artificial Intelligence (AI)/Machine Learning (ML) for NR air interface
3. TR 36.839 Mobility enhancements in heterogeneous networks
4. R2-2403245 Simulation based evaluation of the AIML added mobility Ericsson
5. R2-2402673 Simulation assumption and evaluation methodology NEC
6. R2-2402751 Discussion on simulation assumption and evaulation methodology for AI mobility ZTE
7. R2-2403487 Discussion on simulation assumptions of AI for mobility Nokia, Nokia Shanghai Bell
8. R2-2403498 Discussion on the simulation assumption and evaluation methodology of AI/ML for mobility NTT DOCOMO, INC.
9. R2-2403112 Discussion on simulation assumptions Huawei, HiSilicon discussion
10. R2-2403557 Simulation assumption and evaluation methodology Interdigital Inc.
11. R2-2402406 Simulation Assumption for AI/ML Mobility Intel Corporation
12. R2-2402562 Discussion on Simulation assumption and evaluation methodology vivo
13. R2-2402413 Simulation assumption and evaluation methodology Qualcomm
14. R2-2402445 Simulation Environments for AI/ML-assisted Mobility Samsung
15. 38.901 Study on channel model for frequencies from 0.5 to 100 GHz
16. R2-2402287 Discussion on Evaluation Methodology for AI Mobility MediaTek Inc.
17. R2-2402168 Discussion on RRM measurement prediction OPPO
18. R2-2402589 Discussion on RRM measurement prediction Samsung
19. 38.901 Study on channel model for frequencies from 0.5 to 100 GHz
20. R2-2402748 Discussion on RRM measurement prediction ZTE Corporation
21. R2-2402552 Initial consideration on RRM measurement prediction CMCC
22. R2-2402403 Areas of interest for RRM measurement prediction Intel Corporation

# Annex1 Measurement model



# Annex2 Agreements in RAN2#125bis

**Agreements**

1. For cell level measurement prediction model, at least consider the following cases:

Case 1: To predict beam level results, then generate cell level results based on the predicted beam results;

Case 2: To directly predict cell level results based on cell level results.

Case 3: To directly predict cell level results based on beam level results

1. We will consider intra-frequency intra and inter-cell spatial domain measurement predictions, for beam and cell level measurements.
2. For temporal domain measurement prediction, we will consider the AI-PHY beam management Case A and Case B from the RAN1 AI/ML PHY TR and it applies to both beam level and cell level. As baseline we will focus on pure temporal prediction.
3. The following items can be considered as a baseline for the prediction accuracy of the cell-level measurement prediction：

Spatial-domain prediction： RSRP difference to the actual measurement

Temporal prediction:RSRP difference to the actual measurement

measurement reduction rate as one KPI

1. As a first step we will focus on measurement prediction accuracy. FFS whether and what system level performance evaluation is needed

**Agreements to start evaluations**

* FR1-to-FR1
  + Focus on intra-frequncy in time domain prediction for the purpose of measurement reduction
  + Study inter-frequency scenario in terms of which scenarios can be studied without requiring new channel model and also resolving any simulation assumptions (if possible).
* FR2-to-FR2
  + Focus on intra-frequency
  + Perform evaluation both in time and spatial domain

Agreements

1 AI mobility SI uses synthesized datasets based on 3GPP agreed channel model and deployment for evaluation. Field data is optional

2 Reuse current RAN1’s simulation assumptions as much as possible by extending data generation to neighbouring cells.

3 Once a set of simulation parameters and assumptions per each sub-use case (e.g., propagation scenario, deployment topology, channel modelling, UE trajectories, etc.) are settled, it should be used for baseline case (i.e. without AI/ML model), training (e.g. data set generation), validation, and inference etc.

4 Clarify and document the use of random seeds in between the training and test dataset, simulation drops/runs at least for channel modelling and UE trajectory.

5 Alignment of simulation assumptions is necessary, but explicit result calibration (e.g., as in TR 36.839) is not expected. Companies can independently report their gains achieved by AI/ML with detailed evaluation descriptions for cross-checking purposes.

6 For FR1, band n77/n78 is considered with 4GHz as the central frequency. FFS any other band

7 For FR2, only FR2-1 is considered, e.g., band n257. 30GHz central frequency can be adopted to reuse RAN1’s work as much as possible. FFS any other band

8 focus on Urban Macro (UMa) for FR1 and Umi for FR2

9 RAN2 takes hexagonal regular topology as the starting point.

10 Take baseline simulation assumptions from Table 6.3.1-1 in TR 38.843 for FR2 as the starting point for channel modelling, e.g., BS/UE antenna configuration, BS Tx power, and BS/UE antenna height. UE rotation is excluded in the initial phase of evaluation.

11 UE trajectory model uses options 1-3 in TR 38.843 section 6.3.1 as the starting point. Down-selection to be discussed in email discussion

12 AI/ML model generalization could be addressed after sufficient performance gains for different use cases are found.