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

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

Agenda Item: 8.3.2.1

Source: OPPO(Rapporteur)

Title: Summary of [POST126][031][AIMob] Simulations (OPPO)

Document for: Discussion, Decision

# Introduction

This the summary of following post email discussion:

* [POST126][031][AIMob] Simulations (Oppo)

Intended outcome: Agree to evaluation documentation and small simulation related FFS (needed to start simulation evaluation for August meeting)

Endorse Skeleton TR

Deadline: short

|  |  |  |
| --- | --- | --- |
| Company | Delegate | Email address |
| NTT DOCOMO | Xin Wang | wangx@docomolabs-beijing.com.cn |
| Ericsson | Cecilia Eklöf | cecilia.eklof@ericsson.com |
| Samsung | Taeseop Lee | taeseop.lee@samsung.com |
| vivo | Xiang Pan | panxiang@vivo.com |
| Apple | Sasha Sirotkin | ssirotkin@apple.com |
| CATT | Tangxun | tangxun@catt.cn |
| Xiaomi | Xing Yang | [Yangxing1@xiaomi.com](mailto:Yangxing1@xiaomi.com) |
| Turkcell | İzzet Sağlam | izzet.saglam@turkcell.com.tr |
| ZTE | LiuJing | [liu.jing30@zte.com.cn](mailto:liu.jing30@zte.com.cn) |
| Huawei, HiSilicon | Dawid Koziol | dawid.koziol@huawei.com |
| CMCC | Fang Xie | xiefang@chinamobile.com |
| Nokia | En`drit Dosti | [endrit.dosti@nokia.com](mailto:endrit.dosti@nokia.com) |
| Charter Communications | Phillip Oni | c-phillip.oni@charter.com |

# Discussion

## Simulation report template

It is expected simulation result on RRM measurement use case will be submitted to RAN2#127 meeting for further evaluation after the summer. In order to document simulation results reported by each company, a report template is necessary to be aligned among companies. In RAN2#126, contributions [1] and [2] proposed their understanding of how such a template can be. Table 1 lists the parameters based on agreements made so far.

|  |  |  |  |
| --- | --- | --- | --- |
| Report parameters | | **Company A** | **……** |
| Reported simulation assumptions | UE trajectory option (option 1,2,3 in[4]) |  |  |
| UE trajectory boundary processing option (option 1,2,3 in[4]) |  |  |
| UE speed (30,60,90,120 Km/h) |  |  |
| Inter-frequency correlation assumption in general (yes or no)(Note 1) |  |  |
| Measurement reduction rate(50%,…Note2) |  |  |
| Prediction window (100ms,… Note 3) |  |  |
| Any other parameters (Note 4) |  |  |
| Data Size (Sample number) | Training/validity |  |  |
| Testing |  |  |
| AI/ML model  input/output | Model input (Note 5) |  |  |
| Model output(Note 6) |  |  |
| AI/ML model description | Model type (e.g., LSTM, CNN, transformer …) |  |  |
| Model complexity in a number of parameters(M) |  |  |
| Model complexity in model size (e.g. Mbyte) |  |  |
| Computational complexity [FLOPs] |  |  |
| Metrics | Average L3 cell level RSRP difference (dBm) |  |  |
| Other optional KPIs (e.g., L1 beam level RSRP difference,) |  |  |
| ... | ... |  |  |

Table 1

*Note1: Only applicable for FR1 to FR1 inter-frequency prediction. It should be N/A, if not applicable*

*Note2: Only applicable for intra-frequency prediction, either temporal domain case B or spatial domain. It should be N/A, if not applicable*

*Note3: Only applicable for intra-frequency temporal domain case A. It should be N/A, if not applicable*

*Note4: This could be any other parameter e.g.,* *Inter-frequency shadow fading correction (e.g. full, partial, no),* *Number of configured beams, observation window(ms) etc.*

*Note5: Apart from input of RRM sub case 1,2,3, any other input information e.g. L1 filtering for L1 beam measurement, UE location , , information of input cells are also captured here*

*Note6: Apart from output of RRM sub case 1,2,3, other output e.g. information of output cells is captured here too*

For prediction window, companies seem to be fine to align at least one value. And up to submitted simulation result, it is open for modification in RAN2#127 meeting. During [AT126][030][AIMob] discussion people seems to agree with rapporteur that it should be multiple times of sample period. Considering the FR1 and FR2 channel will be quite different and agreed sample period is also different, we’d better assume different prediction window for them also.

Question 1: What value(s) do you recommend for prediction window for RRM measurement use case for FR1 and FR2 respectively?

|  |  |
| --- | --- |
| Company | comments |
| NTT DOCOMO | We suggest using the value N\*Measurement Period, where the Measurement Period is decided by the following table (Table 3). Considering the time span of the whole HO procedure, we suggest at least considering a long prediction window case to check the capability of AI/ML, e.g., N=5. |
| Ericsson | Agree with DOCOMO (e.g. max value of N=5). |
| Samsung | Considering that the RRM prediction results can be used to prepare the HO in advance, the length of prediction window needs to be aligned with the typical HO preparation time (e.g., 40 ~ 60msec) between source/target gNB. Too long prediction window may need to be considered later.  Our recommendation is   * FR1: 40ms or 80ms (1x or 2x sample period) * FR2: 40ms or 60ms (2x or 3x sample period) |
| vivo | For BM-Case 2 in TR 38.843, the prediction window is 80ms/160ms/320ms/640ms/800ms /others. One straightforward way is down-selecting value(s) among them.  As we already agreed that measurement event prediction can be based on RRM measurement prediction result, the prediction window of RRM prediction should cover the length of TTT. Currently, one typical value of TTT is 320ms. Therefore, 320ms can be baseline for the prediction window and can be used for both FR1 and FR2 evaluation. Besides, there can be multiple prediction results within the prediction window, e.g., every 80ms.  In addition, short time of stay is another KPI that is expected to be optimized with AI/ML-based mobility, whose typical value is 1s. Specifically, if the quality of the target cell is predicted to turn unacceptable after UE completes RACH to the cell, HO to the target cell should not be triggered to avoid the short time of stay or ping-pong handover. Therefore, 1s can be another optional value for the prediction window.  In summary, we propose: For the prediction window, 320ms is the baseline and there can be multiple prediction results within the prediction window, e.g., every 80ms. In addition, 1s can be optional. |
| Apple | Considering this is a very preliminary number anyways, we don’t have a strong view. Having said that, we shouldn’t start from a number which is too high (and 320ms appears too high for us at this sta |
| CATT | As the input for measurement event, the RRM measurement prediction result should cover some of the samples for L3 measurement. Hence, we agree with DOCOMO that value for the prediction window is the value N\*Measurement Period. |
| Xiaomi | We understand it’s related to the targeted goal.  If the goal is to improve HO performance, the prediction window can be aligned with HO preparation time, which can be small, e.g. 40 or 80 ms.  If the goal is to reduce measurement, the prediction window can be aligned with measurement period, e.g. N\* measurement period. |
| Rap | The views are quite diverse ☹. The discussion in section 2.2 suggests that company are fine with measurement period 480ms and 200ms for FR2 and FR1 respectively. Then N\*480ms (or 200ms) is a challenging one. Since RAN2 agreed that prediction window is open to adjust, my suggestion is that we start with less challenging one for FR2 first. How about 5\*20ms=100ms? Note company can still report their prediction window in case they adopt a different value from final aligned one. |
| Turkcell | We agree with Rapporteur |
| ZTE | Agree with DOCOMO, the prediction window should be N\*Measurement Period. Otherwise, the output of the model will not be L3 filtered measurement results. 100ms is too short and it implies the output can only be L1 filtered results, not L3.  In addition, we haven’t discussed the detailed L1 filtering model, but if we assume the UE has 8 Rx beams, then at least 8\*20=160ms will be needed to obtain one round of measurement results per Rx beam.  To ensure the output will be L3 filtered cell results, our recommendation is:   * FR1: 200ms (1\* 200ms) * FR2: 400ms (1\* 400ms);   (for measurement period, we assume that measurement period is 400ms in FR2-to FR2 measurement). |
| Huawei, HiSilicon | Even if the measurement is used to improve handover performance, then predictions after the HO has been already executed are not so useful. We think we should rather consider typical TTT values. Typical values are 160, 320 or 480 ms. We can agree that at least 160 ms should be mandatory and companies may evaluate other values as well to see how prediction window impacts the prediction accuracy. |
| CMCC | Agree with DOCOMO and CATT, the prediction window should cover some samples for L3 measurement. Therefore, the prediction window could be the value N\*Measurement Period (e.g. max value of N=5). |
| NTT DOCOMO 2 | Thanks Rapporteur for the suggestions. We think the 100ms prediction window is too short. For L3 measurement prediction, the basic unit should be the L3 Sample Period, and it is expected that the AI/ML can generate a few predicted samples to enhance the mobility performance. |
| Nokia | We think 100ms could be ok as starting point, but other values should also not be precluded as they might be needed for the other use-cases. As an example, for RLF prediction it may be necessary to align the prediction window with the T310 length and for measurement event prediction with the TTT length. |
| Charter | Agree with DOCOMO and ZTE, prediction window should be N\*Measurement period and the output should be L3 filtered. |

Question 2: Apart from parameters listed in Table 1, what other parameter(s) need be reported? If yes, please provide detail parameter, corresponding description and justification.

|  |  |
| --- | --- |
| Company | comments |
| Ericsson | Historical observation window length (for L1 measurements) for frequency and temporal prediction. It would be good the companies provide the information about the observation window length. It can be defined as the number of samples used as input to the model.  Rap: it is covered in the “any other parameters”. Please check Note4 |
| vivo | 1. applicable condition   Unlike AI beam, AI mobility evaluation is not limited to the same cell, so the model can be a per-cell model (e.g., the training data is collected from UEs in the same cell) or a per-area model (e.g., the training data is collected from UEs in the whole simulation area). For the latter case, to achieve acceptable accuracy, the model size may be quite large.  Therefore, the applicable condition (validity area, e.g., per cell or per area) of the model needs to be reported so that we can fairly compare model performance and model complexity.  Rap: The problem is that RAN2 doesn’t discussed cluster approach sufficiently. But if company really want to provide simulation based on cluster approach, you can put such information in the “model input” and “model output”.  [vivo2]: the applicable condition is related to model generalization and is not coupled with the cluster approach. Specifically, per-cell model can be only applicable to specific cell, and model switch will happen when serving cell changes. While pre-area model can be applicable to the entire simulation area. For both per-cell and per-area models, the input and output can be the same cell, i.e., intra-cell prediction.   1. HO parameter   The handover parameters/handover strategy will have an impact on the distribution of the dataset. For instance, if the A3 Offset is set as a higher value, the UE may experience lower RSRP of the serving cell at the cell edge. To have similar distributions of RSRP, we propose the handover parameters should also be reported or we can just align a set of handover parameters (e.g., HO parameters in TR 36.839).  Rap: But for RRM measurement prediction, does it really matter whether a cell is serving cell or neighbouring cell?  [vivo2]: Can be revisited for the measurement event prediction. |
| Apple | 1. Agree with E/// to report observation window 2. We acknowledge vivo’s comment on cell-specific vs. general models; if a company uses anything but a single model (e.g. multiple cell specific models), this needs to be reported 3. Other than the above too we don’t think anything else is needed |
| Huawei, HiSilicon | 1. As mentioned in our comment above, it would be good to explicitly spell out which model inputs need to be reported and this includes, e.g. observation window length, as mentioned by Ericsson. It is now hidden in the Note as an example and we are concerned that companies may omit this one when providing their results. 2. Another model input which is worth reporting is number of UEs simulated in one drop which gives some general understanding of the generalization of the model. 3. Other than that, we think that we should add an optional metric of “Average L3 beam level RSRP difference (dBm)”. L3 beam level results are used in HO procedure and are part of UE measurement reports, so it is useful to evaluate as well. |
| CMCC | Agree with E/// and Huawei that the observation window needs to be reported for temporal and frequency prediction. |
| Nokia | Agree with other companies that observation window is needed. Suggest to also include the following entries as part of the reporting template:  For intermediate KPI on measurement accuracy:   |  |  | | --- | --- | | General assumptions | | | Parameters | Description | | Measurement event parameters | A3 threshold, hysteresis, time to trigger | | Measurement gap configuration | Measurement gap repetition period, measurement gap length | | … | … |   For system level KPI on mobility performance:   |  |  | | --- | --- | | General assumptions | | | Parameters | Description | | Usage of random seeds – training | Spatial channel model, mobility, etc | | Usage of random seeds – inference | Spatial channel model, mobility, etc | | Measurement event parameters | A3 threshold, hysteresis, time to trigger | | RLF parameters | Qin, Qout, N310, N311, T310 | | RACH parameters | RACH model, execution delay | | Handover parameters | Preparation delay, T304 | | Measurement gap configuration | Measurement gap repetition period, measurement gap length | | … | … | |  |  | |  |  | |
| Charter | Agree with Ericsson. Also, agree with Nokia on using the KPI reporting template. |

Question 3: For parameters in Table 1, any further comments?

|  |  |
| --- | --- |
| Company | comments |
| NTT DOCOMO | For AI/ML input and output entry in Table 1, we suggest adding a note that information about the cluster-based approach, including the numbers of input and output cells and their relations, can be reported there to capture the agreements on the cluster-based approach during the last meeting.  Regarding complexity, we suggest reporting the per-cell values for the cluster-based approach since the per-cell approach may require the model to run multiple times to generate the prediction for all cells concerned. For a fair comparison, the normalized value with respect to the output cell number should be reported.  Rap: I add “information of input cells” and “information of output cells” to cover your comment and vivo’s comments (valid area).  DCM2: Thanks rapporteur for the consideration. We are fine with the suggestions. To echo @Apple and @Huawei’s concern, in our opinion, it can be up to the companies to choose whether to use the per-cell or cluster-based approach as long as it is reported in the table, since this approach has been defined in RAN2. Either approach generates the same predicted measurement results. Therefore, it can be viewed as implementation issue. |
| Ericsson | Agree with DOCOMO. |
| Samsung | Agree with NTT DOCOMO |
| vivo | The current measurement reduction rate is only applicable for intra-frequency prediction. At the last meeting, FR1 to FR1 inter-frequency (frequency domain) is set as high priority for measurement reduction as well.  Therefore, the definition and suggested value of measurement reduction rate for inter-frequency prediction should also be provided.  Rap: Can you clarify what does it mean? To me, inter-frequency prediction means the model will predict a cell of frequency B based on the measurement of co-located cell of frequency A in order to save measurement gap i.e. the reduction rate is fixed.  [vivo2]: From our understanding, inter-frequency prediction may have two approaches:   * Approach 1: Input f1 -> output f2, i.e., UE does not need to perform measurement on f2 at all. * Approach 2: Input f1 (set A+B) + f2 (set B) -> output f2 (Set A), i.e., UE still needs to perform measurement on f2 for set B.   We are OK with the interpretation of Rapp to focus on approach 1. Suggest adding some clarification to align the understanding. |
| Apple | 1. How “model output” is different from “metric”?   Rap: model output is defined by agreed 3 RRM use cases. For sub case 1, it is L1 beam level measurement; for sub case 2 and 3, it is L3 cell level measurement after L3 filtering. Whether it is intra-cell or inter-cell, depends on detail scenario we agreed at last meeting. Metrics refer to performance of the model. The evaluation is based on the output of the model and corresponding label in benchmark case i.e. without AI/ML model.   1. In RAN#126 we have only defined the cluster-based approach, we have not agreed to evaluate it. Therefore, we shall not explicitly mention it. |
| CATT | Agree with DOCOMO. |
| Turkcell | Agree with NTT DOCOMO about reporting the per cell values for the cluster based approach. |
| ZTE | We’d like to further clarify note 4 (any other parameters): We have agreed some simulation assumptions are up to companies. And for some simulation assumptions, we only provide the recommended value, other values are not precluded. These simulation assumptions also need to be reported, include at least:   * Whether LOSsoft is modeled or not; * The number of UE Rx beams; * The number of gNB Tx beams * Other BS/UE Antenna Configuration, BS Tx power, Spatial consistency.   We are also fine to report the information (e.g. number of input cells, per-cell RSRP difference if the number of output cell >1) for cluster-based approach. |
| Huawei, HiSilicon | We have similar understanding about the cluster approach as Apple. Cluster-based approach has not been mentioned in the scenarios prioritization table and in our understanding we will not be evaluating this at least in the initial evaluations phase. |
| Charter | Agree with DOCOMO and no further comment. |

## RRC parameters

Few parameters are left not agreed during [AT126][030][AIMob] discussion as following:

|  |  |
| --- | --- |
| L3 filtering parameter for both FR1 and FR2 | Recommended value |
| FR1 FilterCoefficient | 4 |
| FR2 FilterCoefficient(Note 6) | 4 |

Table 2

|  |  |
| --- | --- |
| Measurement period | Recommended value |
| FR1 to FR1 intra-frequency w.o. gap | 200ms |
| FR1 to FR1 inter-frequency with gap | 200ms |
| FR2 to FR2 intra-frequency w.o. gap | 480ms |

Table 3

|  |  |
| --- | --- |
| Consolidation parameter | Recommended value |
| nrofSS-BlocksToAverage for FR1 | 1 |
| nrofSS-BlocksToAverage for FR2 | 3 |
| absThreshSS-BlocksConsolidation for FR1(Note 7) | -156dbm[2] |
| absThreshSS-BlocksConsolidation for FR2(Note 7) | -156dbm[2] |

Table 4

*Note 6,7: These two parameters are added by rapporteur in case they could be different between FR1 and FR2*

*Note 7: the recommended value from [2] is just for discussion purpose.*

If you have better recommendation, please provide your value(s):

|  |  |
| --- | --- |
| Company | Recommended values |
| NTT DOCOMO | For the measurement period of FR1-to-FR1 inter-frequency with gap (in Table 3), there is no 120ms configuration for the measurement gap repetition period (MGRP) in TS38.331. Although the measurement period does not mean the same value should be used for MGRP, we think it is beneficial for the future study (e.g., monitoring, data collection, etc) if an aligned value can be adopted. Therefore, we suggest using 160ms, which is also closer to the value we used for cases w/o MG.  We are fine with other parameters.  Rap: I thought it makes more sense to align sample period and MGRP, or?  [DCM2] Thank you for your suggestions. We are also fine with 200ms. |
| Ericsson | We are fine with the proposed values. |
| Samsung | -156 dBm of absThreshSS-BlocksConsolidation is too small to measure in our view. The main scenario of this measurement and prediction is mobility-related decision e.g. handover. absThreshSS-BlocksConsolidation should be a typical value indicating the cell could be a serving cell. The exact value could be different between frequencies. But we prefer a common threshold for both FR1 and FR2. Our recommendation of absThreshSS-BlocksConsolidation is -100 dBm or similar value.  Rap: I am bit lost why the predicted cell should be a serving cell. Prediction of neighbouring cell is at least necessary for measurement event prediction when serving cell is also being predicted. If we have such high value, it basically means for neighbouring cell, only top one beam is taken into account. I am wondering whether it is a good approach. On the other hand, -156dbm is the minimum value based on table 10.1.6.1-1 in 38.133. It basically means nrofSS-BlocksToAverage is almost always 3 for FR2. For FR1 it doesn’t make any difference considering so far company are fine with nrofSS-BlocksToAverage=1  [Samsung2]  - We wanted to say that -156 dBm is too small to detect in the real world scenario. This value is the minimum value defined in TS 38.133, but signal below -120 dBm is rarely detected practically, and serving cell RSRP is most likely greater than -100 dBm. The intention of absThreshSS-BlocksConsolidation is beam consolidation among good beams above the threshold. -156 dBm does not represent good beam at all.  We prefer to have a common measurement period if possible. 200ms can be used for both intra- and inter-frequency scenarios. For FR2, we may use the minimum value (400ms), similar to FR1. |
| vivo | From our understanding, a measurement period of 200/480ms means that we will get a L3 filtered measurement result every 200/480ms.  However, the granularity seems a bit large and we think that the measurement period in the simulation should be the same as the L1 sampling period (e.g., 40ms for FR1 and 20ms for FR2). |
| Apple | Agree with Samsung to strive to have a common measurement period. No strong view otherwise. |
| CATT | Regarding the measurement period for FR1 to FR1 inter-frequency with gap, we think it should be derived based on Table 9.3.5-1 “Measurement period for inter-frequency measurements with gaps (Frequency FR1)” in 38.133, instead of Table 9.3.4-3 “Time period for time index detection (Frequency range FR1)” in 38.133 (proposed in [2]). And we suggest using 200ms for this case accordingly.  Rap: I agree. And considering some company want to align between intra-frequency and inter-frequency case for FR1, I think 200ms is reasonable value. |
| Xiaomi | We are fine with proposed values. |
| Turkcell | We prefer to use 200 ms for FR1 to FR1 inter-frequency with gap as CATT suggested. |
| ZTE | **For table 3:**  Since we have agreed that we don’t simulate existence of measurement gap as starting point, no need to discuss measurement period value for FR1-to-FR1 inter-freq measurement with gap now. We are fine to adopt 200ms for both FR1 intra-freq and FR1 inter-freq.  For FR2-to-FR2 intra-freq measurement without gap, based on the table 9.2.5.2-2 of TS 38.133, our recommended value is 400ms.  **For table 4:**  We are fine with proposed value for FR1. But, if *nrofSS-BlocksToAverage* is set to 1, no need to discuss *absThreshSS-BlocksConsolidation* value, since the cell measurement quantity is always derived based on the highest beam measurement quantity value.  For FR2, we tend agree with Samsung that -156dBm is not practical, if the third ranked beam is really bad, it will impact the cell quality, our suggestion is -100dBm or -110dBm. |
| Huawei, HiSilicon | We are fine with the currently proposed values (including using 200ms for both intra- and inter-frequency measurement period in FR1).  Since the number of beams to average is anyway limited by the nrofSS-BlocksToAverage, we have no strong view on the current value of absThreshSS-BlocksConsolidation. |
| CMCC | We are fine with the proposed values. |
| Nokia | We wonder if the values in table 4 need to be aligned (for example the threshold values are very low so likely they do not have significant impact). Tables 1-3 are ok. |
| Charter | We agree with the proposed values as the starting points but sympathize with Samsung that -156 dBm might be too small to be detected. |

## TR skeleton

Please provide your comments directly on TR skeleton [3] in the email discussion folder **without** changing original text.

# Reference

1. R2-2404485, Simulation based evaluation of AIML aided mobility, Ericsson
2. R2-2404713, Discussion on simulation assumption of RRM measurement, OPPO
3. R2-2405693 TR 38.744 Skeleton of AI mobility NR OPPO draft TR Rel-19 38.744 0.0.1 FS\_NR\_AIML\_Mob
4. R2-2405941 Summary of [POST125bis][021][AIML mobility ] Simulation assumptions and methodology OPPO discussion Rel-19 FS\_NR\_AIML\_Mob Late