**3GPP TSG-RAN WG2 #127 *R2-24XXXXX***

**Maastricht Netherlands August 19th – 23th, 2024**

Agenda Item: 8.3.2.1

Source: OPPO

Title: Draft Summary of [POST127][030][AI mobility] RRM simulation assumptions-phase 2(OPPO)

Document for: Discussion, Decision

# Introduction

This short post email discussion intends to discuss the left issue from offline [1].

After the phase 1 of [POST127][030][AI mobility], it is agreed to continue discuss the definition of temporal domain case B and corresponding observation window and prediction window.

# Discussion

## The definition of intra-frequency temporal domain case A and case B

After phase 1 discussion, there are two examples of case B:



Figure 2.1-1 temporal domain case example 1



Figure 2.1-2 temporal domain case example 2

Majority company support example 2 as illustrated in Figure 2.1-2 and some company think example 1 illustrated in Figure 2.1-1 is still valuable. For non-sliding L1/L3 approach, the minimum timing granularity is one measurement period i.e. 200ms for FR1 (400ms for FR2). Assuming the MRRT is 80% then the observation window will become 1200ms for FR1 if there are two actual measurement results are contained within the observation window as illustrated in Figure 2.1-3:



Figure 2.1-3 example2 assuming MRRT=80%

This is one example of FR1. For FR2 the observation window will be simply doubled i.e., 2400ms, which is not necessary long technically in temporal domain. In order to leave some flexibility for non-sliding L1/L3 approach, example 1 as illustrated in Figure 2.1-1 should be allowed in order to simulate some short observation window. There are few more example MRRT, OW and PW in table 1 in Annex\_1 for non-sliding L1/L3 filtering approach for your information.

**Q1: Do you agree to use both example 1 in Figure 2.1-1 and example 2 in Figure 2.1-2 for temporal domain case B?**

|  |  |  |
| --- | --- | --- |
| Company | Yes or no? | comments |
| vivo | example 2 should be the baseline | For the MRRT=80%, the pattern of example ~~2~~ 1 can be as follows:    Our simulation shows that the prediction accuracy will decrease with the increase of the temporal distance between the prediction and the latest measurement. As a result, the prediction accuracy at instance 10 in the above figure will be unacceptable.  [ZTE] In our view, the above figure is actually example 1, not example 2, the difference between example 1 and 2 is whether the OW includes predicted occasions. So for example 2 with MRRT=80%, our understanding is aligned with the figure that provided by email rapp:    For example 1 with MRRT= 80%, it can be:    But we are not so optimistic on the accuracy of MRRT=80% for both example 1 and 2.  [vivo2]: Sorry for the typo. And we don’t think one sample can be input to predict multiple samples in the above example 1 from ZTE.  In addition, we don’t see a technical issue of the observation window length in Figure 2.1-3. Both of these two examples could utilize the same number of historical samples as input. I.e., no additional storage is needed for example 2.  Besides, we already have the same pattern of example 1 for Case A. Since it is a study item, no need to have separate definitions for the same pattern with the same prediction accuracy.  In general, we propose example 2 should be the baseline and example 1 can be optional. |
| Ericsson | No | Example 2 to be used as baseline, since example 1 resembles Case A. Using one measurement-to-prediction pattern would help better comparison of the results. |
| NTT DOCOMO | No | Agree with vivo and Ericsson that Example 2 is used as the baseline. |
| CATT | Yes | Actually, we don’t think the long observation window is one big issue for AI simulation considering AI model mainly learns the measurement result in sequence and tries to predict the skipped measurement result in Case B.  But it’s ok for us that companies can choose which example is used and report in the result table. |
| ZTE | Yes | As replied to Vivo’s comments, we think the main difference between example 1 and example 2 is whether there are predicted occasions in OW.  We think both example 1 and example 2 are useful use cases for temporal domain prediction case B. Although example 1 looks similar to Case A, but they indeed have different performance impact if the predicted occasions will never be measured by the UE.  Especially for non-sliding window, example 1 is more realisitic, because for example 2 with high MRRT, the measured samples within OW will be very far away and OW will be very large which results in low prediction accuracy.  It will be good if companies have the same understanding on example 1 and example 2. |
| MTK | Yes | We agree the pattern of case B is left to the companies’ design, which includes example 1, example 2, and other options. Companies can provide their result with the corresponding illustration of their pattern. Given that, under the same MRRT, OW, and PW, the number of measured/predicted results in OW/PW may vary for different patterns, and we know that the number may impact the AI performance. Therefore, we encourage companies to specify the number of measured/predicted results in OW/PW. |
| Huawei, HiSilicon | OK, but see comments | We think example 2 should be the baseline. We agree with vivo that example 1 of Case B is very likely to perform worse than example 2 with the same reduction rate due to big distance between the measured instance and predicted instances. But we are also OK to allow the companies simulate example 1 if they want to, but the companies should of course clarify the pattern they used together with the results. Also, it seems the main difference between Case A and Case B example 1 is that in the latter the predicted measurement instances are never truly measured. In Case A on the other hand, we do not perform any measurement reduction. We could make it clear in the definition. |
| Nokia | Yes | Same view as MTK, that this can be left for companies to choose. Agree that reporting the pattern (including whether predicted samples were used as an input or not) is needed. |
| Interdigital | See comment | Example 2 should be the baseline |
| Xiaomi | Yes | We think both can reach the goal of measurement reduction. Another benefit of example 1 is UE can obtain the ground truth L3 fitered measurement result as model input. In example 2, the UE can’t obtain the ground truth L3 filtered results since UE skipped measurement every other time instance. |

There is a key word in definition of temporal domain case B :

**Intra-frequency temporal domain case B-To be discussed in Phase 2:**

In case B, measurement results in prediction window are predicted by historical measurement result(s) in observation window. Then observation window and prediction window slide forward with either sampling period(s) (with sliding L1/L3 filtering option) or measurement period(s) (with non-sliding L1/L3 filtering option) after measurement result(s) in previous prediction window is/are skipped.

In the example illustrated in Figure 2.1-2, the “grey” block is previous predicted measurement result(s) but in OW. The issue is whether those predicted measurement results should be input to model. At this stage majority company who already use this example pattern doesn’t take predicted measurement result into account. Maybe we can assume those predicted measurement results are not used. But company can also report if they use it.

**Q2: The historical measurement results in OW are actual measurement results. And company can report it, if they use predicted measurement results (if any) in OW as input of model also. Do you agree with this proposal?**

|  |  |  |
| --- | --- | --- |
| Company | Yes or no? | comments |
| vivo | Yes | If the previous prediction is used as model input, the uncertainty of previous prediction may have an impact on the subsequent prediction. Therefore, we are not sure whether this input could increase the prediction accuracy and we are OK to study. |
| Ericsson | See comment | We think predictions (in particular if inferred with a low accuracy) as input to the model may lead to a worse performance and may create unpredictable behaviour in results that makes the comparison of the results more difficut.  Therefore we suggest as baseline it is better to include only the measurements in the OW. We think having one approach would ease the comparison of the results. |
| NTT DOCOMO | See comment | We think this issue is up to the implementation for either the UE-sided or NW-sided models. Looping back the model output may be beneficial for some types of AI/ML model structures, but it may not be for others. It can be viewed as an inner structure of the AI/ML models. It is not necessary to mention it explicitly. |
| CATT | Yes | But we have the sympathy that the predicated results used as input may bring impacts on the prediction accuracy. |
| ZTE | Yes | We are fine to leave it to company to report whether predicted results are used as model inputs.  In fact, from simulation perspective, it is hard to completely preclude the predicted samples. For instance, for use case2 (L3 cell ->L3 cell), there are several ways to generate the input data:    Option 1: L3 cell result at T2 is filtered based on the [real] L3 cell results obtained at T1;  Option 2: L3 cell result T2 is filtered based on the L3 cell results obtained at T0;  Option 3: No L3 filtering for L3 cell results at T2.  We can also leave it to company to report which option is used for generating the input data in case of use case 2. |
| MTK | Yes | We think the case with actual measurements can be the baseline. The study of using predicted measurement can be optional. |
| Huawei, HiSilicon | See comments | Agree with Ericsson and MTK that using real measurements as an input should be the baseline. But if companies would like to also check how including predicted results as an input impacts the performance, it is OK. In any case, what is used as an input needs to be reported by the companies. |
| Nokia | Yes | Using predicted measurement results as an input also affects the complexity of the model, so companies should report whether it’s being used as input or not. |
| Interdigital | See comments | Upto companies to consider using earlier predicted measurements as input for further predictions, as long as it is reported in the results. |
| Xiaomi | NO | Agree with Ericsson that only ground truth measurement result shall be used as model input. The prediction inaccuracy would be accumulated. |

## Observation and predication window

For temporal domain case B, even for the example 2 as illustrated in Figure 2.1-2 there could be more variable way to interpret it e.g. whether prediction window can be prolonged further. So, it becomes very difficult to list all the possible patterns and their combinations for company to select. The simulation results submitted to RAN2#127 meeting show what matters for prediction accuracy of temporal domain case B are following key parameters:

1, MRRT

2, Observation window length

3, Prediction window length

So, a relatively simple way to align these 3 parameters is to set up a value range for them, and here are recommended value ranges from rapporteur:

* MRRT: {50%~80%}
* OW: {200ms ~ 2000ms}
* PW: {40ms~800ms}

Company can report their MRRT and corresponding OW and PW when providing simulation result. In addition, in order to show the potential impact by detail pattern, company can also report their detail pattern.

**Question 4: Please provide your comment on the way to align the simulation assumption for temporal domain case B for FR1:**

|  |  |
| --- | --- |
| Company | comments |
| vivo | Agree with rapp to align the 3 parameters and company should report the detailed pattern of both input and output. |
| Ericsson | Agree |
| NTT DOCOMO | Agree with Rapp’s comments. |
| CATT | Agree |
| ZTE | Agree |
| MTK | We agree with considering aligned ranges for MRRT, OW, and PW, but we think for the sliding/FR1 case, both minimum OW and PW can be 20ms (the sampling period is 20ms and Rx =1) |
| Huawei, HiSilicon | Agree with Mediatek that the minimum value of observation window can be shorter, but we think it can be regardless of L1 filtering option. |
| Nokia | Agree to report MRRT, OW, PW, and detailed pattern. |
| Interdigital | Agree with Rapp’s comments, but it would be good to have at least one or two combinations of parameters agreed upon as mandatory/baseline so that we are sure we have results that can be comparable across all companies (e.g., take one or two rows from the examples given in the Annex). |
| Xiaomi | Fine with the proposal |

# Conclusion

# Reference

[1] R2-2407781 Summary of [AT127][026][AI Mob] Simulation assumptions (OPPO) OPPO discussion

[2] R2-2497849 Summary of [POST127][030][AI mobility] RRM simulation assumptions (OPPO)-Phase 1

# Annex\_1

|  |  |  |  |
| --- | --- | --- | --- |
| MRRT | OW1 | OW2 | PW |
| 50% | 200 | 1000 | 200 |
| 67% | 200 | 800 | 400 |
| 75% | 200 | 1000 | 600 |
| 80% | 200 | 1200 | 800 |

Table 1 examples of OW, PW for FR1 for non-sliding L1/L3 filtering approach

In table 1, the OW and PW is calculated by repeating a pattern few times by following the way illustrated in Figure 2.1-2. For example, a pattern is one measurement result is predicted/skipped after one actual measurement result and it is repeated 3 times. The MRRT is thus 50% and OW=1000ms, PW=200ms i.e., the green row in Table 1 and illustrated in Figure 2.1-3.