**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 (OPPO)

Document for: Discussion, Decision

# Introduction

During RAN2#127 meeting, after discussing offline summary [1], RAN2 concluded that:

**Agreements**

1 To keep two filtering options on the table and up to company to report.

*2* One fixed sampling period of FR2 is introduced for L1/L3 filtering option 1 to replace existing one i.e., 20ms. The detail value is FFS.

3 In the definition of 3 RRM sub-cases, all cell level measurement result(s) refers to L3 filtered cell level measurement

*4 continue to discuss following issues in the post email discussion:*

*1, Further clarification of intra-frequency of temporal domain case A and case B.*

*2, The set of observation vs prediction window parameters for intra-frequency temporal domain case A and case B*

*3, The number of TX and RX for FR1 and FR2*

*4, Filtering co-efficient for beam level prediction*

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

# Discussion

## The definition of L1/L3 filtering option

During offline discussion [1] RAN2 confirmed following two observations:

Observation 1: filtered L1 and L3 RSRP are produced per sample period in option 1

Observation 2: filtered L1 and L3 RSRP are produced per measurement period in option 2

The L1/L3 filtering options will be mentioned in the context of this email discussion. And it is likely they need be captured in TR also. So, first suggestion from rapporteur is that proper definition is needed at least for the sake of discussion. To make it bit easy, let’s call them sliding L1/L3 filtering option i.e., option 1 and non-sliding L1/L3 filtering option.

**In sliding L1/L3 filtering (i.e. option 1), filtered L1 or L3 RSRP are generated every sample period**

**In non-sliding L1/L3 filtering (i.e. option 2), filtered L1 or L3 RSRP are generated every measurement period**

*Note: Filtered L1 RSRP refers to filtered L1 beam level measurement result. And filtered L3 RSRP refer to filtered beam or cell level measurement result.*

**Question 1: Do you agree with above definition of sliding L1/L3 filtering and non-sliding filtering option?**

|  |  |  |
| --- | --- | --- |
| Company | Yes or no? | comments |
| NTT DOCOMO | Yes |  |
| Huawei, HiSilicon | The definition is incomplete | We suggest the following clarifications for both options 1 and 2:   1. In both options, the filtered L1 measurement result is obtained using the non-filtered L1 measurement results obtained during the time duration equal to a measurement period 2. In both options, filtered L3 measurement result is obtained as specified in section 5.5.3.2 of TS 38.331.   It would be actually good to capture some figures as well and the ones from the offline discussion summary were good in general. The only clarification that would have to be added there is that there is L3 filtering happening to derive L3 filtered result which was missing. |
| OPPO | see | Fine to capture the Figure 2.1-1 and Figure 2.1-2 in @meeting offline summary R2-2407781 in the TR. And also agree 2 clarifications from Huawei are also correct observations. |
| MTK | Yes | We agree with the naming and definition. |
| Nokia | Yes | We think the proposed definitions suffice for simulation purposes. |
| Ericsson | Yes |  |
| Xiaomi | Yes |  |
| CATT | Yes |  |
| Samsung | Yes with comment | In general, we share the view with HW. We can focus on L1 filtering (not L3 filtering) in the definition. L3 filtering method is already specified in 5.5.3.2 in TS 38.331. I.e., a new L3 filtered value is calculated every time there is a receiving measurement results (L1 filtered value) from the physical layer. |
| CMCC | Yes with comments | Agree to use the name “sliding L1/L3 filtering” and “non-sliding L1/L3 filtering” for L1/L3 filtering option 1 and option 2, respectively. And agree the clarifications for both options from Huawei. |
| ZTE | Yes with comments | For Option 1, suggest to further clarify the meaning of sampling period in FR2 scenario, such as:  **In sliding L1/L3 filtering (i.e. option 1), filtered L1 or L3 RSRP are generated every sample period (or every sample period \*Rx beams in FR2)**  Note: it can be updated based on the outcome of Q3. |

Summary:

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

First of all, we need clarify what is temporal domain case A and case B. And what does observation window (OW) and prediction window (PW) mean for case A and case B.

For both case A and case B, the observation window means the window covering historical L1 or L3 RSRP to predict future L1 or L3 RSRP, while prediction window means the window covering future L1 or L3 RSRP to be predicted.

In case A, once the measurement results in prediction window are predicted, the observation window and prediction window move forward with either one sampling period (L1/L3 filtering option 1) or one measurement period (L1/L3 filtering option 2).



Figure 2.2-1

For temporal domain case B, the detail pattern could still be flexible. What is agreed at RAN2#127 meeting:

**Agreements**

=> companies are encouraged to considers both prediction from low-frequency cell to high-frequency cell and prediction from high-frequency cell to low-frequency cell, but only low to high is expected.

=> For the agreed frequencies for inter-frequence case, only one UE speed is considered for inter-frequency prediction in simulation, e.g., 30km/h. Companies can consider other speeds for other frequencies if they chose to simulate them.

=> **For temporal domain case B prediction the input is historical measurement values and the output is the values at the subsequent time instances that measurement is skipped, i.e., the prediction is always after the measurement and is at future time instance(s).**

Table 2.2-1

Here are two examples of temporal domain case B:



Figure 2.2-2



Figure 2.2-3

In Figure 2.2-2 i.e. example 1, when the measurement result(s) in prediction window are predicted, both observation window and prediction window move forward as such that the measurement in previous prediction window is skipped. In Figure 2.2-3 i.e. example 2, the observation window and prediction window work in the same way as example 1. The only difference is that some historical measurement results are reused for prediction operation. So, the essential difference between temporal domain case A and temporal case B is whether the predicted measurement results are actually measured or skipped. If they are measured that’s case A, case B otherwise.

**Observation 3: The essential difference between temporal domain case A and temporal case B is whether the predicted measurement results are actually measured or skipped. If they are measured that’s case A, case B otherwise**

Here is recommended definition from rapporteur based on the discussion above:

**Intra-frequency temporal domain case A:**

In case A, 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 one sampling period (with sliding L1/L3 filtering option) or measurement period (with non-sliding L1/L3 filtering option) after measurement result(s) in one more sampling period (with sliding L1/L3 filtering option) or measurement period (with non-sliding L1/L3 filtering option) is(are) actually measured.

**Intra-frequency temporal domain case B:**

In case A, 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.

*Note: the change mark in case B reflects the difference between these case A and case B*

**Question 2: Do you agree with above definition of intra-frequency temporal domain case A and case B? If no, please elaborate.**

|  |  |  |
| --- | --- | --- |
| Company | Yes or no? | comments |
| NTT DOCOMO | No | We agree that Figure 2.2-1 is Case A, and Figure 2.2-3 is Case B.  As to Figure 2.2-2, it is actually a hybrid version of Case A and Case B. It reduces the measurement overhead, which corresponds to study goal #1. Meanwhile, it also has a strong ability to predict over a long prediction window, which is the same as Figure 2.2-1 and can be used to improve the mobility performance corresponding to our study goal #2.  At the current stage, we prefer to focus our study on Figures 2.2-1 and 2.2-3. We can study the hybrid case later if necessary. |
| Huawei, HiSilicon | Case A is OK  Case B – see comments | We suggest to clarify that in Case B prediction and observation windows can be interlaced. Based on this, we suggest the following definition:  “In case B, measurement results in prediction window are predicted by historical measurement result(s) in observation window. In Case B, some of the measurement results in the prediction window may be actual measured results while others are predicted results, depending on the applied prediction pattern. Also, depending on the prediction pattern, the observation and prediction window slide forwards with either one or more sampling period(s) (with sliding L1/L3 filtering option) or with one or more measurement period(s) (with non-sliding L1/L3 filtering option) after each prediction. |
| OPPO | Yes | To Docomo: if the pattern in Figure 2.2-2 is used to save measurement, the length of observation and prediction window supposed to be much shorter compared to those in Figure 2.2-1 because otherwise the prediction accuracy will become worse.  To Huawei: not sure about the statement “In Case B, some of the measurement results in the prediction window may be actual measured results while others are predicted results, depending on the applied prediction pattern”. If there is an actually measured result, it is historical one, so this part should be part of observation window instead of prediction window, right? If I misunderstand, maybe you can provide a Figure to illustrate your point? |
| MTK | see comment | We agree that one of key differences between cases A and B is whether the target predicted time slot will be measured or skipped. For case A, we agree the observation window(OW) contains multiple continuous measure results and prediction window (PW) contains multiple continuous predicted results immediately following the OW as shown in Figure 2.2-1. However, the device does not necessarily perform inference in every sample period. Other options, e.g., slide forward with multiple steps, are also allowed as shown in the following figure.    We recommend revising the description “In case A, ***continuous*** measurement results in prediction window are predicted by historical ***continuous*** measurement result in observation window. Then observation window and prediction window slide forward with either ***~~one~~*** sampling period***(s)*** (with sliding L1/L3 filtering option) or measurement period***(s)***,…”  For case B, since RAN2 allow companies to consider different patterns under the same measurement reduction ratio, the OW/PW may not necessarily contain continuous measured/predicted results. Also, in different patterns, the same OW/PW may contain different numbers of measured/predicted results. For example, assume the sampling interval is 20ms and OW is 80ms, For the pattern in Figure 2.2-2, the OW contains 4 measured results, but in Figure 2.2-3, the OW only contains 2 measured results. Therefore, for case B, we recommend companies also provide the number of measured/predicted results in OW/PW under their pattern.  We recommend revising the description, “In case B, measurement results in prediction window are predicted by historical measurement result(s) in observation window. ***The number of measurement results in observation/prediction window depends on the pattern and is reported by companies.*** Then observation…” |
| Nokia | Yes, but see comments | It’s not clear what the “historical measurement results” in Case B Example 2 refer to. Do they include predicted measurements, or are only confined to measured ones? It would be good to clarify, because reusing predicted measurements as an input has an impact on performance. |
| Ericsson | Yes, but see comments | For Case B we think that the Example 1 resembles Case A and for the sake of alignment among different companies maybe we could agree to use only the pattern illustrated in Example 2 for the Case B. Limiting the simulation assumptions to Example 2 for Case B helps having a like-for-like comparison of the results provided by the companies. |
| Xiaomi | Yes |  |
| CATT | Case A is OK. Case B with comments. | For case B, we share the similar view that Example 2 in Figure 2.2-3 is our understanding, and Example 1 is same as case A. |
| Samsung | Yes | Regarding the comments from Docomo/HW, we have the same understanding with OPPO. |
| CMCC | Case A is OK. Case B with comments. | Similar view with CATT. |
| ZTE | See comments | For case A:   1. The term “sampling period” may need to be clarified, considering 20ms\*N Rx beams in FR2 scenario.   For case B:  We think both Example 1 and Example 2 can be considered in simulation. Both of them can achieve the goal of measurement reduction. Although Example 1 looks similar to case A, they have different HO performance impact if we plan to do system-level simulation;  For Example 2 of case B, there are two issues need to be clarified:   1. As Nokia and MTK commented, whether the observation window only includes measured samples, or also includes predicted samples?      1. If the observation window only includes measured samples, then it depends on what values of OW/PW will be used. For example, in both sliding and non-sliding L3 cases, at least [3N, N] should be used. [N, N] is not applicable because it is same as Example 1. [2N, N] and [4N, N] are also not applicable because they are the same as [N, N] and [3N, N]. Similarly, [N, 2N] and [N, 4N] are not applicable, because odd positions will be actual measured samples. Please correct me if I misunderstood anything. |
|  |  |  |

## Observation and predication window

RAN2 agreed “One fixed sampling period of FR2 is introduced for L1/L3 filtering option 1 to replace existing one i.e., 20ms. The detail value is FFS”. The intention to have one fixed value is to decouple the FR2 L1 sampling period from the number of RX and leave the L1 sampling among RXs to company’s implementation.

Rapporteur recommend the new value is 80ms, but company want to check internally.

**Question 3: Do you agree 80ms as sampling period for FR2? If not, please indicate your preferred value**

|  |  |  |
| --- | --- | --- |
| Company | Yes or no? | comments |
| NTT DOCOMO | Yes |  |
| Huawei, HiSilicon | See comments | Not sure why we need to have a single fixed value? Couldn’t it depend on the number of employed Rx chains, i.e. 20xN, where N is the number of Rx chains? |
| OPPO | Yes | We think more detail discussion about how to do L1 sampling for FR2 can be avoided if we just have one value, which is also help to compare simulation among companies. Plus, the agreed RX number for FR2 is N=4 and 20\*4=80ms 😊 |
| MTK | See comment | We agree with Huawei’s comment. We can set the sampling period as 20\*k ms directly where k is the number of Rx beams. |
| Nokia | Yes |  |
| Ericsson | Yes | We agree with OPPO comment and for the sake of alignment on the simulation assumptions we prefer to use a fixed value i.e., 80ms. |
| Xiaomi | Comments | Agree with HW |
| CATT | Comments | Agree with HW |
| Samsung | Comments | Agree with MTK. The real sampling rate depends on the assumed # of Rx beams which is up to companies. |
| CMCC | Comments | Agree with HW |
| ZTE | Comments | Tend to agree with HW, but if we are able to agree “4 Rx beams” for FR2 in Q5, 80ms will be ok. |

Summary:

The prediction window and observation window should be the multiple times sample period or measurement period. Based on observation 1 and 2, they are different between L1/L3 filtering options. In addition, based on the submitted simulation results from company, it seems the ratio between observation window and prediction window matters. Here are examples of the potential observation window vs prediction window:

|  |  |  |
| --- | --- | --- |
| OW vs PW | Sliding L1/L3 filtering option | Non-sliding filtering option |
| Ratio | N=80,120,160,200 (Note 1) | N=400,800,1200 |
| 3N:N |  |  |
| 2N:N |  |  |
| N:N |  |  |

Table 2.3-1 example of observation window vs prediction for FR2 temporal domain case A

Note 1: the value is related to new sampling period in **Q3** and the minimum step should be the new value. They should be updated once Q2 is answered.

|  |  |  |
| --- | --- | --- |
| OW vs PW | Sliding L1/L3 filtering option | Non-sliding filtering option |
| Ratio | N=80,120,160,200 | N=200,400,600 |
| 4N:N |  |  |
| 3N:N |  |  |
| 2N:N |  |  |
| N:N |  |  |
| N:2N |  |  |
| N:3N |  |  |
| N:4N |  |  |

Table 2.3-2 example of observation window vs prediction for FR1 temporal domain case B

In table 2.3-1 and 2.3-2, by putting N into the ratio, the length of OW and PW are obtained. For example N=400 for ratio 3N:N, it means OW vs PW is 1200ms vs 400ms.

**Question 4: Do you agree set of OW and PW in table 2.3-1 and 2.3-2? If not, please indicate your preferred value**

|  |  |  |
| --- | --- | --- |
| Company | Yes or no? | comments |
| NTT DOCOMO | No | Multiple historical values should be observed for prediction accuracy. We do not think the OW with length N or 2N makes sense. A longer OW is expected to improve prediction performance. Some AI/ML models, such as LSTM, can memorize and utilize all the historical measurements without an explicitly defined OW.  We suggest that the OW be equal to or longer than the PW and include the longer OW cases, such as 5N. Then, the ratios may be,  5N:(1-5)N  4N:(1-4)N  Or we consider the following 3 ratios to simplify the issue,  5N:5N  4N:4N  3N:3N |
| Huawei, HiSilicon | Yes, but see comments | In our understanding the companies are free to choose a subset of these options and are not expected to check all of them. Also, we are not so sure about the cases where the observation window is much longer than the prediction window (e.g. 4N:N). From our experience too big OW/PW ratio increases calculation complexity, but does not bring much gains.  On the other hand, we are OK to let companies consider larger N values as suggested by Docomo. |
| OPPO | See comments | We think to have a limited pool could help compare company’s simulation result.  To Docomo: your proposed 3 new ratio is for which table 2.3-1, right? For table 2.3-2 a longer prediction window will make the predication accuracy worse even observation window is equally long. |
| MTK | No | It is unclear why we have different N values since we already agreed that the sampling period is 20/40ms(or 20\*k/40\*k if considering Rx beam based on Q3) and the measurement period is 200/400ms for FR1/FR2.  Also, we think the combinations are too many, we can just consider 4N:N, N:N, N:4N for both cases A and B. Other options are not precluded but are not mandatory. |
| Nokia | See comments | Similar view as DCM, HW and MTK. It can be difficult to align on a single value in this case. Considering much larger OW/PW can have benefits in some cases (not always though), but overall would result in higher complexity. We think we can consider 4N:N, N and N:4N cases, and companies can additionally present results for which they see benefits. |
| Ericsson | Yes, but see comment | We don’t see any reason that the prediction window should be limited to N in Case A while in Case B the prediction window is extended to 4N. We believe given the goal of Case A is to enhance the handover performance, the UE should be able to predict the radio link quality further in the future e.g., 2N, 3N, 4N. Therefore, we suggest Table 2.3-1 should include longer prediction window (e.g. N:2N, N:3N, N:4N) as in Case B.  In addition, if 80 ms is agreeable to be used as sampling period for FR2, we think the other options should be removed from the Table 2.3-1. |
| Xiaomi | Comments | In general, it’s fine. However, it’s related to whether 80ms is agreeable. N can be other values. |
| CATT | Comments | For FR2 in Table 2.3-1, we think the predication window can be even larger, considering TTT and the reporting delay in case of helpful event A3 prediction. Therefore, at least 2N:2N and 3N:3N can be considered. |
| Samsung | Yes with comment | In general, we share the view with HW. Companies can choose some sets of OW/PW in the table for their evaluation. It should not be mandatory to evaluate all the cases in the table.  For table 2.3-1, we propose to consider the following options since we observe that the OW shorter than PW can show reasonable accuracy with less input data size in Case A prediction.  N:2N  N:3N |
| CMCC | Comments | More ratios of OW/PW could be considered in evaluation, e.g., 4N:N, 5N:N for FR2 temporal domain case A. |
| ZTE | See comments | The variable N and variable [xN, yN] results in many options. It is better to assume a fixed N value for non-sliding and sliding cases. For example:   * For temporal domain case A, N = 80ms for sliding case; N=400ms for non-sliding case; * For temporal domain case B, N=40ms for sliding case; N=200ms for non-sliding case.   Then we can further discuss different [xN, yN] options.  We think the discussion does not preclude companies to consider other OW/PW values, but it is better to at least agree a minimum set of values. In our understanding, it could be:   * For case A:   --[2N, N], [3N, N], [4N, N] (to evaluate the impact caused by different OWs)  --[2N, 2N], [3N, 3N], [4N, 4N] (to evaluate how far AI can predict)  Note: For sliding option, since smaller N is used, it is better to consider larger PW in order to compare with non-sliding case, e.g. [10N, 10N]   * For case B:   --[2N, 2N], [3N, 3N], [4N, 4N] (to evaluate 50% reduction rate)  --[2N, 3N], [2N, 4N] (to evaluate 60%, 66.7% reduction rate) (OW=N may not be reasonable because only one value is used as the input)  Note: it also depends on which example (1 or 2) is used as we commented in Q2. |
|  |  |  |

## TX and RX numbers

From the contributions of this meeting, for FR2 here are the current chosen values:

TX: {8,12,32,64}, RX {1,4,8}.

For FR1 here are the current chosen values:

TX: {1,4,8,12,32}, RX {1,2,4}.

**Question 5: Which RX and/or TX number do you prefer to keep for FR1 and FR2 respectively?**

|  |  |  |
| --- | --- | --- |
| Company | Yes or no? | comments |
| NTT DOCOMO |  | There are some ambiguities between the antenna port and antenna element numbers for the numbers listed here. We prefer the format in our agreed simulation assumption tables, i.e., (M, N, P, Mg, Ng, Mp, Np), which makes it clearer to set up the simulations.  For FR1, we prefer (8,8,2,1,1,2,8), which has 32 ports and 128 antenna elements for Tx, and (1,1,2,1,1,1,1), which has 2 ports/antenna elements for Rx.  For FR2, we prefer (4,8,2,1,1,1,1), which has 2 ports and 64 antenna elements for Tx, and (1,4,2,1,2,1,1), which has 4 ports and 8 antenna elements per panel (16 antenna elements in total) for Rx.  These configurations have been agreed in RAN2 #126. |
| Huawei, HiSilicon |  | We agree with Docomo that the companies should simply use what we had already agreed before.    C:\Users\y00781912\AppData\Roaming\eSpace_Desktop\UserData\y00781912\imagefiles\D0859BD0-ED55-4755-B460-8B2BF7305A55.png  From the options agreed before, our preference is: For FR1, TX: 32, RX: 4  For FR2, TX: 2, RX: 4 |
| OPPO |  | Let’s stick to the agreed one |
| MTK |  | For simulation comparison, we tend to agree one Tx/Rx beam number as the starting point, where we prefer  FR1: Tx 1, Rx 1  FR2: Tx 32, Rx 4  Different Tx/Rx beam numbers are not precluded if companies would like to provide their observation from different settings. |
| Nokia |  | Agree with HW and Oppo that we can re-use the values that were agreed before. If we should choose one value, we prefer to select the smallest values (as they would decrease the simulation runtime). |
| Ericsson |  | We prefer:  FR1: Tx 4, Rx 1  FR2: Tx 8, Rx 1 |
| Xiaomi |  | Agree with Ericsson |
| CATT |  | We also would like to stick to the original BS/UE Antenna configuration.  And we prefer:  For FR1, Tx = 4, Rx = 1;  For FR2, Tx = 32, Rx = 4. |
| Samsung |  | We would like to clarify the intention of the original question. If the Rx/Tx means the # of Rx/Tx antenna port, we support to reuse the agreed configuration (as indicated by HW’s comment). On the other hands, if the Rx/Tx means the # of Rx/Tx beams, we think it can be up to companies based on the agreement. |
| CMCC |  | We prefer:  For FR1, Tx = 4, Rx = 1;  For FR2, Tx = 32, Rx = 4. |
| ZTE |  | We think the Rx/Tx means Rx/Tx beams. For FR2, the UE is supposed to have at least 4 Rx beams. So, we prefer:  For FR1, Tx = 4, Rx =1  For FR2, Tx = 32, Rx =4 |

## Filtering co-efficient for Beam level prediction

Filtering co-efficient for cell level prediction is 4 i.e. k=4. The simple way is to reuse it for beam level prediction.

**Question 5: Do you agree the filtering co-efficient i.e. parameter k=4? If not, please provide your preferred value.**

|  |  |  |
| --- | --- | --- |
| Company | Yes or no? | comments |
| NTT DOCOMO | Yes |  |
| Huawei, HiSilicon | Yes | It can be aligned with the assumptions for cell level L3 filter. |
| OPPO | Yes |  |
| MTK | Yes | We can reuse k=4 for beam level prediction |
| Nokia | Yes |  |
| Ericsson | Yes | K=4 |
| Xiaomi | Yes |  |
| CATT | Yes |  |
| Samsung | Yes |  |
| CMCC | Yes |  |
| ZTE | Yes | K=4 |

# Conclusion

# Reference

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