**3GPP TSG-RAN WG2 Meeting #127bis**  **R2-240xxxx**

**Hefei, China, Oct 14th~ Oct 18th, 2024**

Agenda Item: 8.3.2

Source: Mediatek Inc.

Title: [POST127bis][016][AI Mob] Simulation results (Mediatek)

Document for: Discussion, Decision

# Introduction

This report provides a summary for the following post-meeting email discussion:

* [POST127bis][016][AI Mob] Simulation results (Mediatek)

Intended outcome: finalize the table (one week email deadline)

- Informational email discussion on logistics of storing simulation results.

The deadline of the email discussion is Oct. 25th, 10:00 UTC. Please provide your comment by Oct. 25th, 8:00 UTC to allow us sufficient time to revise the table.

Companies' comments on the spreadsheet examples from the at-meeting email discussion [AT127bis][016][AI Mob] Simulation Table Example (Mediatek) are attached in Appendix 1 for your reference. **To save time, you do not need to repeat comments that were already expressed and addressed in the at-meeting email discussion. However, you are welcome to provide additional comments or suggestions for refining the templates if your previous comments were not properly addressed.**

Companies providing input to this email discussion are requested to leave contact information below.

|  |  |  |
| --- | --- | --- |
| **Company** | **Name** | **Email Address** |
| ZTE | Song Xiaohui | song.xiaohui@zte.com.cn |
| OPPO | Hao Wu | wuhao8@oppo.com |
| Apple | Sasha Sirotkin | ssirotkin@apple.com |
| NTT DOCOMO | Xin Wang | wangx@docomolabs-beijing.com.cn |
| Ericsson | Cecilia Eklöf | cecilia.eklof@ericsson.com |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

# Discussion

The offline discussion will collect companies' comments and suggestions on the example spreadsheets based on current agreements. Any further discussion beyond what has been agreed upon is not within the scope of this discussion.

Please notice the following revisions on the template:

1. New columns have been added according to the newly reached agreements, with the content written in red.
2. An example row has been added to define the format of each table's content.

## Comments for the spreadsheet examples (Scenario 2, Scenario 4, Scenario 3, Scenario 6)

Please provide comments on the spreadsheet examples for the following scenairos:

* Example1\_Scenario 2: RRM Measurement Prediction Evaluation results for intra-frequency temporal domain case B;
* Example 2\_Scenario 4: RRM Measurement Prediction Evaluation results for intra-frequency temporal domain case A;
* Example 3\_Scenario 3: RRM Measurement Prediction Evaluation results for inter-frequency (frequency domain);
* Example 4\_Scenario 6: RRM Measurement Prediction Evaluation results for intra-frequency spatial domain.

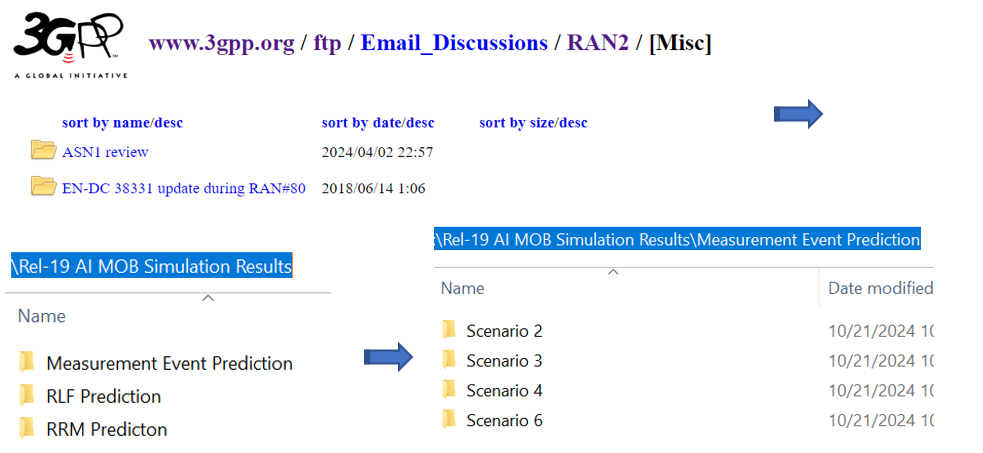
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Company | Example1\_Scenario 2 | Example 2\_Scenario 4 | Example 3\_Scenario 3 | Example 4\_Scenario 6 |
| Huawei | We have the same comments as during the AT-meeting offline, i.e.:   1. Clarify that the dataset size is expressed in the number of samples (some companies expressed it in bytes previously, so it is better to avoid confusion), e.g. “Training data size in number of samples [K]”   [Rapp] Will revise according to your comment.   1. Non-AI method should be treated as just another kind of model, so it seems natural that the results are provided in separate rows and the model description can contain additional info about the utilized method. Comparison between AIML and non-AIML can be easily done in companies Tdocs, similarly as we did, e.g. for different UE speeds.   [Rapp] I have addressed your and Apples comments in the updated spreadsheets. Please check whether you have further suggestions. | | | |
| Apple | 1. Same comment as Huawei 2), but I see it has been already addressed in the example spreadsheets 2. I assume “Number of cell” (in Model Related Information) refers to “Number of cell**S**”, please clarify   [Rapp] Yes. Number of cells, which can be one or multiple. Will update.   1. Furthermore, if it is inded “number of cells”, how would it work in case of model outputs? What’s the meaning of a “single cell level RSRP difference” in this case? I guess this is for cluster approach with multiple outputs (is it?), in which case it would be rather hard to capture in such a spreadhsheet. | | | |
| Apple | 1. I suppose “L3 cell RSRP difference (dB)” refers to “average RSRP value over the entire window”, so it needs to be clarified (i.e. add the word “average”).   [Rapp] You are right. But we have provided the definition in the KPI spreadsheet. Please check the description/definition, ‘Average of RSRP difference between predicted and actual L3 cell-level measurement result at all measurement points within the PW’. | |  |  |
| NTT DOCOMO | 1. We also feel that RSRP differences for multiple cells should be further clarified. We are studying the mobility enhancements, so the final predictions should involve multiple cells (including serving and neighbouring cells) whenever cell-specific or cluster-based models are used. For both approaches, the definition of the RSRP differences should be clarified:    1. If cell-specific models are used, clarify whether the reported RSRP differences are for the serving cell, one or more neighbouring cells, the Top-1/K cell at the prediction time, or averaged across cells (after running the models for each cell).    2. If cluster-based models are used, clarify whether the reported RSRP differences are for the Top-1/K cell at the prediction time or averaged across cells’ corresponding model outputs. 2. The definition of the complexity for the cluster-based approach also needs clarification. For the cluster-based approach, one inference can obtain the prediction of multiple cells, while for the cell-specific approach, the model should be run multiple times to obtain the same outputs. Should a normalized complexity (by cell number) be reported for the cluster-based approach? | | | |
| Mediatek | Comment to Apple 3) and Docomo 1) for RSRP difference.  Agree with NTT Docomo. The issue is common to both cell-based and cluster-based approaches. For RRM measurement, the UE needs to perform RRM measurements for both the serving cell and neighboring cells. For RRM prediction, it makes no sense to perform RRM prediction only for the serving cell while conducting actual measurements for neighboring cells. Therefore, L3 cell RSRPs for both the serving cell and neighboring cells need to be predicted, implying that the L3 cell-level RSRP differences for multiple cells need to be considered for accuracy performance evaluation. The difference between the cluster-based and cell-based approaches lies in whether a single AI model or multiple AI model inference operations are performed.  My suggestion is to clarify how to handle multiple cells in the upcoming RAN2 meeting, and I will add an FFS issue for the KPI definition. The following options can be considered. Companies please report in their Tdoc which method are being used.   * Multiple L3 cell-level RSRP difference values (e.g., Top K cells with the best ground-truth values) * Average of the L3 cell-level RSRP difference values cross multiple cells * L3 cell-level RSRP difference of the serving cell only * Other? | | | |
| OPPO | We agree that non-AI results can be provided in separate rows, as is currently done. Previously, we have also agreed that both non-AI and simple AI methods can be used for comparison. Therefore, it is better to clarify whether an AI model applied is used as a “baseline” (i.e., simple AI without neural networks) or as an AI enhancement method. One simple way is to label them separately as “simple AI” and “AI”. | | | |
| Ericsson | We think it would be clearer if “L3 cell RSRP difference” could be changed to “Average L3 cell RSRP difference” in the “KPI” sheet, to highlight that it is an average of avarages in the PW. | In the “KPI” sheet: same comment as in Example 1.  Not clear what the purpose is of “L3 beam RSRP difference(dB)” in this scenario. | OK | OK |
| ZTE | OK | OK | Although we have agreed to report correlation coefficient, since there are several ways to get the correlation coefficient, it seems a little early to determine to use PCC. It’s unclear of how to calculate and use PCC. We observe that in the [POST127bis][022][AI mobility] Simulation Assumptions, OPPO provides a question (Question 28) to collect companies’ view for the PCC, we suggest to keep it as FFS, and update it until agreement is reached. | OK |
|  |  |  |  |  |

## Logistics of Storing Simulation Results

During this meeting, we agreed on how to facilitate companies in providing simulation results for AI/ML mobility.

|  |
| --- |
| **Agreement**:  Create a folder on the 3GPP FTP server for companies to upload their simulation results. Within this folder, create individual subfolders for different use cases. For each use case different subfolders under the main directory corresponds to the different identified scenairos.The file name of the excel table follows the format: 'MeetingNumber\_CompanyName\_TdocNumber\_version number'. |

After consulting with the secretary on the appropriate location to store the simulation results, we plan to create a folder named "Rel-19 AI MOB Simulation Results" within the [Misc]/ directory under the RAN2 Email Discussion folder ([https://www.3gpp.org/ftp/Email\_Discussions/RAN2/[Misc]](https://www.3gpp.org/ftp/Email_Discussions/RAN2/%5bMisc%5d)). Within this folder, we will have subfolders for different use cases and scenarios. This way, all simulation results for each use case/scenario from different meetings will be collected in one place. The structure will be as follows: [Misc]/[Rel-19 AI MOB Simulation Results]/[Use case]/[Scenario]. We will create the folder once we confirm the logistics of storing the simulation results in this email discussion.



|  |  |
| --- | --- |
| Company | Comments/Suggestions |
| Apple | 1. Just to remind that as far as I recall we haven’t yet agreed on scenarios for other use cases (RLF prediction, event prediction, etc). But its ok, the term “scenario” is generic enough. Which reminds me to point out that we have been using the term “use case” for different things, but I guess we all understand what is meant here.   [Rapp] The scenarios for event prediction and RLF prediction have not been agreed upon yet. Therefore, the main directories for RLF Prediction and Measurement Event Prediction are currently blank.   1. What’s the meaning of “tdocnumber” in the filename? We agreed on this in the meeting, but now it is not clear to me what tdoc this refers to. I guess it would be the tdoc where the results are analized? But then, for the November meeting, most companies will submit also the results which have been submitted before. Do we then refer (using the tdoc number) to the cintribution the results were described originally? Or a new one (where we would need to describe the results once again)… Buttom line, I’m not sure the tdoc number is needed. If it is going to be used, we need to clarify what tdoc the number should refer to.   [Rapp] My intention is that when companies have new simulation results to submit, they should provide a Tdoc to summarize the analysis and observations. We will begin collecting simulation results from the November meeting. I assume that companies will (re-)submit or update the simulation results according to the agreed template and submit a Tdoc where the results are analyzed in the upcoming meeting. Therefore, the Tdoc number will start from the November meeting.  In subsequent meetings, companies do not need to resubmit the same results but can update or submit new results based on the discussions. In this case, I assume companies will provide additional analysis for the new results or indicate in their Tdoc that some results submitted in the previous meeting need to be updated. The new Tdoc number information should then be provided. The Tdoc number information also helps us easily track the trends observed by different companies. |
| Ericsson | Agree |
|  |  |
|  |  |
|  |  |

# Conclusion

Proposal：

# Appendix 1- [AT127bis][016][AI Mob] Simulation table example (Mediatek)

The offline discussion will collect companies' comments and suggestions on the example spreadsheets based on current agreements. Any further discussion beyond what has been agreed upon is not within the scope of this discussion.

Please notice the following revisions on the template:

1. New columns have been added according to the newly reached agreements, with the content written in red.
2. An example row has been added to define the format of each table's content.

## Scenario 2

Please provide comments on the spreadsheet example for Scenario 2: RRM Measurement Prediction Evaluation results for caseB in the table below.

|  |  |
| --- | --- |
| Company | Comment/suggestion |
| ZTE | Regarding performance metrics:   1. Based on the agreement, companies can provide multiple real time RSRP value(s), an example is needed for this case. For example, how to show multiple values, whether companies also need to fill in the column ‘Last predicted point L3 cell RSRP difference’?   [Rapp] One colume to provide multiple RSRP values is added.   |  |  |  | | --- | --- | --- | | **Performance Metrics** | | | | **L3 cell RSRP difference  (dB)** | **Last predicted point L3 cell RSRP difference (dB)** | **Multiple predicted point L3 cell RSRP differences (dB)** | | **XX** | **XX** | **[XX, XX, XX..]** |  1. The name ‘Last predicted point L3 cell RSRP difference’ is unclear. Actually, it is also an average value, to average all last predicted point in the prediction window. We can use ‘Average RSRP difference for last predicted point within prediction window’   [Rapp] It’s too long to capture the exact meaning for the metrics in the cell. We can refer the definition in the sheet of KPI.   |  |  | | --- | --- | | L3 cell RSRP difference (dB) | Average of RSRP difference between predicted and actual L3 cell-level measurement result at all measurement points within the PW | | Last predicted point L3 cell RSRP difference (dB) | Average of RSRP difference between predicted and actual L3 cell-level measurement result at the last point within the PW | | Multiple predicted point L3 cell RSRP differences (dB) | Average of RSRP difference between predicted and actual L3 cell-level measurement result for each individual point within the PW. Multiple values are provided if there are multiple points within the PW. |  1. For non-AI, considering it is optional for companies to report, maybe we can mark it as optional.  |  |  |  |  | | --- | --- | --- | --- | | **Performance Metrics** | | | | | **Average L3 cell RSRP difference (dB)** | **Average RSRP difference for last predicted point within prediction window** | **Average L3 cell RSRP difference(non-AI/simple AI) (dB) (optional)** | **Average RSRP difference for last predicted point within prediction window (non-AI/simple AI) (dB) (optional)** | | **XX** | **XX** | **(non-AI/simple AI) Method: XX** | **(non-AI/simple AI) Method:XX** | | 0.198 |  | (non-AI) Sample and hold: 0.198 |  |   [Rapp] This is the study phase, so we don't need to define the evaluation exercises as mandatory or optional. I believe not all metrics are required to be provided by each company. If a company can provide the value, the cell should be filled. Otherwise, it can be left blank. |
| OPPO | 1. It is not clear what coarse update in “Spatial consistency” is. We do not have any discussion about it and it can be company implementation. Purely choice between A and B is good enough. Same comment to other scenarios.   [Rapp] Yes. Revised.   1. The template for “user earlier predicted results as input or not” should be “NO” rather than “No” according to what is highlighted in row 3.   [Rapp] Yes. Revised.  3. For non-AI and simple AI methods, the current agreement is that companies are free to report. Now, it seems to be mandatory. Companies may have different methods and even for sample and hold there could be different implementations. It could be hard for us to get some common observations from those not aligned methods. Therefore, we think it is premature to include them in spreadsheets. The comment also applies to other scenarios.  [Rapp] This is the study phase, so we don't need to define the evaluation exercises as mandatory or optional. I believe not all metrics are required to be provided by each company. If a company can provide the value, the cell should be filled. Otherwise, it can be left blank.  However, to address your concern, we will keep it separate from the performance metrics to avoid any confusion.   |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | | **Performance Metrics** | | | **Non-AI/Simple AI** | | | | **L3 cell RSRP difference  (dB)** | **Last predicted point L3 cell RSRP difference (dB)** | **Multiple predicted point L3 cell RSRP differences (dB)** | **Non-AI/Simple AI Method** | **L3 cell RSRP difference  (dB)** | **Last predicted point L3 cell RSRP difference (dB)** | |
| Huawei | Most comments are common for all scenarios:   1. Agree with OPPO there is no need to have specific “performance metrics” for non-AIML. These results can be provided separately, i.e. by indicating that the model used was a non-AIML model. Otherwise, we will have to indicate two models in one row which will make it messy. The comparison between AIML and non-AIML can be done by companies in their Tdocs. Also, we should stop using “simple” AIML model term. Whether a model is simple or not can be determined based on its parameters (size, FLOPs etc.).   [Rapp] I have split them into two categories, which I hope addresses your concern. As for the term 'simple,' I added it based on our agreement. It may be too late to find a proper name before we reconvene. I can keep it in brackets, indicating that the term can be changed if a more suitable name is found later.   1. For L1 filtering another option is that there is no L1 filtering, so it should be: “(none/sliding/non-sliding)”   [Rapp] I believe this point relates to the type of data used as model input: L1 RSRP (no L1 filtering), L1-filtered RSRP, or L3-filtered RSRP. For all sub-use cases, we need L3 filtering to derive the L3 cell-level RSRP, as described in Figure 5.1-1/2 of the TR. Therefore, I will keep the option as it is, but companies can indicate this e.g. L1 RSRP in model input column.  If no filtering is used, you can leave the cell blank.   1. For the detailed pattern, it is not clear to me what, e.g. “2” means – should we have some examples?   [Rapp]’1/2’ refers to the examples captured in the TR. There is no space to capture the detailed pattern in the table. Companies can indicate which example is referred and provide the detailed example in their Tdoc if necessary.    Figure 5.2.1-2 Example 1 of intra-frequency temporal domain case B    Figure 5.2.1-3 Example 2 of intra-frequency temporal domain case B   1. Training/testing data size – it should be clarified what this means. I think we refer to number of samples, so it should be made clear.   [Rapp] Yes, it refers to the number of samples. We now have a common understanding. Later, we can add more definitions or notes in the spreadsheets/TR if any terms are unclear.   1. RAN2 agreed to also optionally check beam level RSRP prediction accuracy. It should be added to the table as optional metric to report.   [Rapp] We agreed that L3 filtered beam level prediction sub-use cases are lower priority and should focus on FR2 intra-frequency temporal domain case A. I can add one matrics for L3 beam-level RSRP in case A only. |
|  |  |
|  |  |
|  |  |
|  |  |

## Scenario 4

Please provide comments on the spreadsheet example for Scenario 4: RRM Measurement Prediction Evaluation results for caseA.

|  |  |
| --- | --- |
| Company | Comment/suggestion |
| ZTE | Same as comment above |
| OPPO | It is better to remove “1 cell, L3 cell level RSRP ~~in OW~~” in “Model input” and “Model output” to align examples among all scenarios.  [Rapp]: Revised. |
| vivo | The performance metrics should be RSRP difference   |  |  |  |  | | --- | --- | --- | --- | | **Performance Metrics** | | | | | **Average L3-RSRP(dB)** | **Last Predicted Point L3 cell RSRP(dB)** | **Average L3-RSRP(non-AI/simple AI)(dB)** | **Last Predicted Point L3 cell RSRP(non-AI/simple AI)(dB)** |   [Rapp] Revised. |
|  |  |
|  |  |
|  |  |
|  |  |

## Scenario 3

Please provide comments on the spreadsheet example for Scenario 3: RRM Measurement Prediction Evaluation results for frequency.

|  |  |
| --- | --- |
| Company | Comment/suggestion |
| ZTE | Maybe we can mark non-AI column as optional |
| OPPO | 1. For “Channel correlation coefficient between two frequency layers”, there are different correlation coefficient, such as PCC (Pearson Correlation Coefficient), Spearman Correlation Coefficient and Kendall Correlation Coefficient. Given that PCC is simple for use, we can state it clearly that the coefficient is PCC to reduce the ambiguity   [Rapp] Revised.  2. We found that some companies have reported below assumptions in the “other factors” column. How about capturing them to be separate columns?   |  | | --- | | Inter-frequency correlation assumption in general (yes or no) | | Inter-frequency shadow fading correction (e.g. full, partial, no) | |  |   [Rapp] Revised. |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

## Scenario 6

Please provide comments on the spreadsheet example for Scenario 6: RRM Measurement Prediction Evaluation results for spatial.

|  |  |
| --- | --- |
| Company | Comment/suggestion |
| ZTE | Maybe we can mark non-AI column as optional |
| OPPO | Similar to other scenarios, having another main group to reflect the performance of non-AI or simple AI models would be better. |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

## General Rules for Filling the Table

The rules for filling out the table are as follows:

1. Please adhere to the format provided in the example as much as possible. Certain columns, such as "Other Factors, details of AI model" do not have strict content restrictions.
2. Please make sure to keep the same parameter units as the template provided.

Companies are encouraged to provide additional rules to facilitate the recording of simulation results.

|  |  |
| --- | --- |
| Company | Comment/suggestion |
| OPPO | For ease of data analysis, we propose that each blank can only be filled with one value. For example, only one value can be reported in “Last predicted point L3 cell RSRP difference (dB)”. We’d better avoid reporting a set in it, e.g., [0.2, 0.34, 0.56]. |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

**Rules for filling out the table:**

1. Please adhere to the format provided in the example as closely as possible, except for the specified columns. The columns 'Other Factors,' 'AI Model Type,' 'Details of AI Model,' and 'Non-AI/Simple AI Method' do not have strict content restrictions."
2. Please make sure to keep the same parameter units as the template provided.
3. Companies are not required to fill in all the information, especially the metrics. If companies can’t provide the information, please leave the cell blank.

# Appendix 2-RAN2 Agreement in RAN2#127bis

* *For intra-frequency temporal domain, higher UE speeds result in larger prediction errors*
* *Initially, increasing the OW length can enhance prediction accuracy in the temporal domain case A, especially when the OW is relatively short. However, once the OW exceeds a certain threshold, further increases do not yield significant benefits. Conversely, for PW, longer durations correlate with decreased prediction accuracy. RAN2 will not define the actual threshold and fast fading assumption.*
* Majority of companies observe that among sub cases 1, 2, and 3, at least with shorter prediction window sub case 2 demonstrates the highest prediction accuracy
* Companies can provide multiple real time RSRP value(s) and/or average RSRP value over the entire window and should indicate in their simulation results what they have used. The companies should at least provide the results of only one value it should be the last value at the end of the PW. We will add two columns in the spreadsheet to capture the last value and the average value.
* Companies need to report whether earlier predicted results are also used as inputs for future RRM prediction.
* Companies should report with their simulation the correlation coefficient

* Higher-to-lower and lower-to-higher frequency prediction is comparable

For co-located scenario, the UE speed in the inter-frequency case has minor impact on

* prediction accuracy
* Companies are free to consider non-AI or simple AI models

**Agreements**

1 It is mandatory to follow the following rules for filling out the table for simulation results:

* Adhere to the format provided in the example, except for the specified columns. The columns 'Other Factors,' 'AI Model Type,' 'Details of AI Model,' and 'Non-AI/Simple AI Method' do not have strict content restrictions.
* Keep the same parameter units as the template provided.
* Companies are not required to fill in all the information, e.g. some performance metrics. If companies can’t provide the information, please leave the cell blank.

NOTE: The rapporteur will not include the inputs if these rules are not followed

1. Adopt the agreed spreadsheet (after email discussion) examples of different RRM prediction scenarios to capture companies’s simulation results.