**3GPP TSG-RAN WG2 Meeting #129 R2-250xxxx**

**Orlando, USA, Oct 18th – 22nd, 2024**

Agenda item:

Source: Apple (moderator)

Title: Email discussion summary for  [Post128][018][AI Mob] generalization (Apple)

Document for: Discussion and Decision

# 1 Introduction

This is the draft document (and the future email discussion summary) for the following email discussion.

* [Post128][018][AI Mob] generalization (Apple)

Intended outcome: Discuss parameters for different cell configuration and attempt to prioritize 1 parameters and not more than 2 values per parameter. Can do 2 max values if really reneed. for

Deadline: 3 weeks (i.e. December 13)

For your convenience, below you can find the relevant agreements on model generalization from RAN2#128.

|  |
| --- |
| **Agreements on generalization**   1. Reuse the evaluation methodology in TR38.843 for generalization study, i.e., the generalization performance is evaluated with the following cases,  * *Baseline:* The AI/ML model is trained using the dataset with Configuration #B and tested using the dataset with Configuration #B. * *Generalization Case #1 (GC#1):* The AI/ML model is trained using the dataset with Configuration #A but tested using the dataset with Configuration #B. * *Generalization Case #2 (GC#2):* The AI/ML model is trained using mixed datasets with both configurations and tested using the dataset with Configuration #B.   2 Companies can choose which case they compare with and should report it with simulation results.  3 Generalization issues on RRM measurement prediction are prioritized.  4 Start the study with generalization issue with RRM measurement prediction in temporal domain. Companies can chose to study frequency domain prediction cases and report what they have simulated.  5 Study generalization over UE speeds  6 The simulation assumption of FR1 temporal domain case B is reused for generalization study with 3 UE speeds i.e. 30Km/h, 60Km/h and 90Km/h. FFS on combinations  7 The simulation assumption of FR2 temporal domain case A is reused for generalization study with 3 UE speeds i.e. 60Km/h, 90Km/h and 120Km/h. FFS on combinations |

**Further agreements on generalization**

1. Companies that would like to study inter-frequency generalization can start with input 2GHz and output 4GHz, and 4GHz to 2GHz. FFS if we introduce a third frequency.
2. Study model generalization across different cell configurations (e.g. ISD, gNB height, power, beam pattern, etc). FFS which parameters we prioritize.

# 2 Discussion

The following is hopefully a complete list of all the proposals relevant to the present discussion made in contributions submitted to RAN2#128:

* R2-2409652
  + Proposal 3: Study AI/ML generalization from cell perspective for intra-frequency RRM measurement prediction, i.e., whether the model trained based on one cell’s data can also be applied to another cell’s intra-frequency RRM measurement prediction.
* R2-2409668
  + Proposal 1: The following factors are considered in the generalization study for RRM prediction:
    - Scenarios, including deployment scenarios (e.g., UMa, UMi), ISD (e.g., 500m, 200m), UE speed (e.g., 30/90 km/h for goal 1 and 60/120 km/h for goal 2), gNB height (e.g., 25m, 10m), and UE height (e.g., 1m, 1.5m).
    - Configurations (parameters and settings), including gNB settings (e.g., DL Tx beam codebook), UE parameters (e.g., Rx beam number), set B of beam (pairs) for spatial domain prediction, and sample period/measurement period.
* R2-2409829
  + Proposal 5: RAN2 to study generalization issue on RRM measurement prediction in frequency domain with different frequency combinations.
  + Proposal 6: RAN2 to verify the impact of including source/target frequency information as input of AI/ML to study generalization issue on RRM measurement prediction in frequency domain with different frequency combinations
* R2-2409869
  + Proposal 14: RAN2 use the definition of model generalization:
    - Model generalization, i.e., using one model that is generalizable to different scenarios/configurations/areas.
  + Proposal 16: RAN2 consider to study the impact of setting(scenarios/configurations/areas) on model generalization, including UE speed, beam pattern, and the collection area of data.
  + Proposal 18: To evaluate the impact of data collection area, considering following options:
    - Model trained with dataset of one certain cell, apply inference on certain cell (one model for one cell).
    - Model trained with dataset of cluster(all cells), apply inference on any one cell of the cluster (one model for cluster);
    - Model trained with dataset of one cell, apply inference on any other of cells within the cluster (one model for cluster);
* R2-2409972
  + Proposal 7: model generalization refers to generalization across cells with potentially different configuration. It is evaluated in a limited number of scenarios of high priority for RRM measurement prediction.
  + Proposal 9: to use field data for mode generalization study; to discuss simulation assumptions to generate an environment with multiple different cell configurations (for model generalization study).
* R2-2410023
  + Proposal 3
    - Study the generalization aspects according to the prediction domain of the use cases.
      * For Case 2 and 4 (temporal domain), the aspects are
        + UE speed,
        + Temporal domain configurations (measurement period, observation window length, prediction window length).
      * For Case 3 (frequency domain), the aspect is
        + The carrier frequencies and frequency gap between the two bands.
      * For Case 6 (spatial domain), the aspects are
        + TXRU mapping, tilt angle, and SSB/CSI-RS beam number and pattern.
* R2-2410263
  + Proposal 3 For generalization purposes, the following parameters can be prioritized in the study:
    - UE speed: the AI/ML model is trained using a certain UE speed and the AI/ML model performs inference for a UE with a different speed, including stationary UEs.
    - Frequency: The AI/ML model is trained using a certain frequency, but the inference is done using another frequency.
    - Cell size: the AI/ML model is trained for small cells, but the inference is done for large cells.
  + Proposal 8 For generalization evaluation over cell size:
    - Training dataset is small cells, and inference dataset is large cells.
    - Training dataset is large cells and inference dataset is small cells.
* R2-241054
  + Proposal 4: To verify how well the AIML model can be generalized, RAN2 should check the following (a prioritization discussion is expected):
    - how well the model trained in one frequency (e.g. 2GHz) performs in another one (e.g. 4GHz)
    - how well the model trained in a certain deployment scenario performs on another one, e.g. UMa, UMi deployments, different ISDs
    - how well a model trained in a certain cell configuration perform in another one, e.g. gNB/UE antenna heights, different UE/gNB port settings, different number of beams etc.
    - how well the same model works for different prediction window lengths
* R2-2410800
  + Proposal 3: For generalization performance verification, consider the following scenario/configuration:
    - Scenarios:
      * Various deployment scenarios, e.g., UMa, UMi and others; e.g., 200m ISD or 500m ISD and others; e.g., same deployment, different cells with different configuration/assumption; e.g., gNB height and UE height;
      * Various UE mobility, e.g., 30km/h, 60km/h, 90km/h, and 120km/h.
    - Configurations (parameters and settings):
      * Various UE parameters, e.g., number of UE Rx beams, UE antenna configuration, measurement period, L3 filtering parameter, system bandwidth
      * Various gNB settings, e.g., the number of BS Tx beams, BS Antenna configuration.
* R2-2410345
  + Proposal 2: For intra-frequency temporal prediction, the following aspects can be considered to verify the generalization performance of an AI/ML model over various configurations:
    - UE speed (e.g. 30km/h, 60km/h, 90km/h, etc.)
    - the number of TX/RX (e.g. 1/1,1/2,1/4 for FR1, 8/4,16/4,32/4 for FR2)
    - MRRT for FR1 (e.g. 50%, 80%)
    - OW/PW (e.g. 5/1, 4/1, 1/1, 1/2, etc.)

Based on the proposals listed above, it appears that the following would be a fair summary of all the cell configuration parameters proposed for the generalization study (with comments from the moderator):

1. deployment scenarios/channel model

* Moderator’s comments:
  + Only UMa and UMi have been agreed so far
  + Furthermore, UMa was agreed for FR1 and UMi for FR2

1. ISD

* Moderator’s comments:
  + Only 500m and 200m have been agreed so far
  + Furthermore, 500m was agreed for FR1 and 200m for FR2

1. BS antenna height (e.g., 25m, 10m)

* Moderator’s comments:
  + Only 10m and 25m have been agreed so far
  + Furthermore, 25m has been agreed for FR1 and 10m for FR2

1. BS antenna configuration

* Moderator’s comments:
  + In currently agreed spreadsheets, the following is used for FR1
    - 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
  + In currently agreed spreadsheets, the following is used for FR2
    - Antenna setup and port layouts at gNB: (4, 8, 2, 1, 1, 1, 1), (dV, dH) = (0.5, 0.5) λ

1. BS antenna radiation pattern

* Moderator’s comments:
  + In currently agreed spreadsheets, the following is used for FR1
    - 3-sector antenna radiation pattern, 8 dBi
  + In currently agreed spreadsheets, the following is used for FR2
    - TR 38.802 Table A.2.1-6

1. BS antenna tilt

* Moderator’s comments:
  + No explicit agreements in RAN2

1. Number of Tx beams

* Moderator’s comments:
  + 1,2 and 4 have been agreed for FR1
  + 8, 16, and 32 have been agreed for FR2

1. BS Tx power

* Moderator’s comments:
  + This is my interpretation of the proposal to generalize across “small vs large cells”
  + In currently agreed spreadsheets 44 dBm is used for FR1 and 40 dBm for FR2

1. Usage of field data:

* Moderator’s comments:
  + This is not strictly speaking a network configuration parameter, but it sort of incorporates all of them

### Question 1

**Question 1: which of the parameters listed above you prefer to use for the generalization across different cell configurations study? Please try to limit your response to the lowest number of parameters you can accept. Please consider providing additional technical details for your favourite parameters in the comments column, if needed.**

|  |  |  |
| --- | --- | --- |
| Company | Preferred parameters | Comments |
| Ericsson | a), b), c) and h) all together | We believe for a realistic evaluation of the generalization across “small vs large cell”, the parameters a), b), c), h) should be jointly considered. That is, in a real scenario, a change of one parameter implies changes of the other parameters. In this respect we propose the following configurations.   * Only one frequency is used, e.g.: FR1 * Parameter values used for small cell   + a) deployment scenarios: UMi   + b) ISD = 200m   + c) BS antenna height = 10m   + h) BS Tx power = 40dBm * Parameter values used for large cell   + a) deployment scenarios: UMa   + b) ISD = 500m   + c) BS antenna height = 25m   + h) BS Tx power = 44dBm   The “small vs large cell” generalization is done by training the model using the parameters with values for small cell, and inference on the data collected in the network with cells the parameters associated to the large cell (or vice versa). |
| Nokia | See comments | In our view, the order of importance would be:  a) deployment scenarios/channel model, including changing the LOS/NLOS and shadow fading random seeds and UE trajectories  b) ISD and c) BS antenna height  h) BS Tx power  g) Number of Tx beams  Then d) e) f). |
| Huawei | b), c), d), h), all together | We agree with the approach proposed by Ericsson, i.e. to evaluate the model in two different cell configurations, combining several different parameter settings. But in our view, it is better to focus on the UMa deployment with different cell sizes/settings, because such cells are more likely co-exist and being deployed next to each other on a certain area. Also, this would limit the workload a bit as, so far, we considered UMa only. Hence, we propose the following scenarios, starting from Ericsson’s proposal:   * Frequency – FR1 * Parameter values used for Parameter-Set A   + b) ISD = 200m   + c) BS antenna height = 10m   + d) BS antenna configuration 16 ports: (8,4,2,1,1,2,4), (dH,dV) = (0.5, 0.8)λ   + h) BS Tx power = 40dBm * Parameter values used for Parameter-Set B   + b) ISD = 500m   + c) BS antenna height = 25m   + d) BS antenna configuration 32 ports: (8,8,2,1,1,2,8), (dH,dV) = (0.5, 0.8)λ   + h) BS Tx power = 44dBm |
| NTT DOCOMO | g, a, b, c | We have a similar view to Ericsson’s. These parameters are mutually impacted in practical deployments. Therefore, it is not necessary to study them separately. Several combinations can be considered, e.g., considering the following two combinations for FR2,  Config. 1: (g = 32, a = UMi, b = 200m, c = 10m)  Config. 2: (g = 16, a = UMa, b = 500m, c = 25m)  For FR1, different *g) Number of Tx beams* can be further considered based on the two sets of parameters suggested by Ericsson. |
| vivo | Combination of a), b), c) and h) for both FR1 and FR2 as 1st priority and  d) or g) as 2nd priority. | To relieve the workload of simulation, we agree with Ericsson on combining multiple parameters for generalization simulation as a starting point.  However, as the generalization study on FR1 is for temporal domain case B and the study on FR2 is for temporal domain case A, we think both FR1 and FR2 should be considered. If one single FR is to be selected, we think FR2 should be prioritized to optimize mobility performance.  Based on 38.901, UMa and UMi are two typical high-level descriptions of deployment scenarios, which will reflect multiple typical evaluation parameters as follows. For instance, ISD is 200m for UMi and 500m for UMa; BS antenna height is 10m for UMi and 25m for UMa.  Table 7.2-1: Evaluation parameters for UMi-street canyon and UMa scenarios   |  |  |  |  | | --- | --- | --- | --- | | Parameters | | UMi - street canyon | UMa | | Cell layout | | Hexagonal grid, 19 micro sites, 3 sectors per site (ISD = 200m) | Hexagonal grid, 19 macro sites, 3 sectors per site (ISD = 500m) | | BS antenna height | | 10m | 25m | | UT location | Outdoor/indoor | Outdoor and indoor | Outdoor and indoor | | LOS/NLOS | LOS and NLOS | LOS and NLOS | | Height | Same as 3D-UMi in TR36.873 | Same as 3D-UMa in TR36.873 | | Indoor UT ratio | | 80% | 80% | | UT mobility (horizontal plane only) | | 3km/h | 3km/h | | Min. BS - UT distance (2D) | | 10m | 35m | | UT distribution (horizontal) | | Uniform | Uniform |   Therefore, no need to consider the ISD and antenna height individually. Besides, based on the generalization simulation results of beam management in 38.843, with ISD 200m/ISD 500m, for generalization Case 2 compared to Case 1, evaluation results from 3 sources show about 1%~2% degradation.  To study the comprehensive generalization ability of the model in different scenarios for different goals, two sets of parameters in these two scenarios for FR1 and FR2 can be considered. The parameters proposed by Ericsson can be the baseline:   * ~~Only one frequency is used, e.g.: FR1~~ * Parameter values set 1   + a) deployment scenarios: UMi   + b) ISD = 200m   + c) BS antenna height = 10m   + h) BS Tx power = 40dBm * Parameter values set 2   + a) deployment scenarios: UMa   + b) ISD = 500m   + c) BS antenna height = 25m   + h) BS Tx power = 44dBm   As to the impact of gNB antenna configuration, i.e., d), proposed by Huawei, we think it should be studied individually since gNB antenna configuration can be different for the same deployment scenario in different areas. Due to limited time, it can be optional and with 2nd priority. For FR2, as only one antenna setup and port layouts were agreed, the number of Tx beams, i.e., parameter g), can be considered for generalization simulation. |
| ZTE | a), b), c), h) all together | Agree with Ericsson, i.e. to study two different cell configuration set. In our understanding, a) b) c) are not independent, we cannot just study one of them. And for frequency range, we suggest to focus on FR1, since 500m ISD may cause weak coverage in FR2 scenario. |
| Mediatek | a), b), c) and h) all together as UMa/UMi setting  g) Number of Tx beams | Agree with Ericsson's approach to jointly consider the generalization method for a), b), c) and h). RAN2 has agreed that UMa will be used for FR1 and UMi for FR2. When evaluating UMa/UMi, a set of settings including ISD, BS antenna height, and BS Tx power should be considered.  We think field data may be useful for the model generalization study as ultimate generalization test. However, it is challenging to have a consensus on the field dataset. We are open to verifying the performance of the AI approaches via field data if company (for example, Apple, if possible) could provide the field dataset as a common dataset. |
| Samsung | See comments | In general, we have some doubts on whether the generalization study with different cell configuration “in simulation environment” is really useful. Anyway the uniform/symmetric cell topology/configurations in simulation are unrealistic and far from the real field environment. In that sense, we share the view with MTK that the generalization study via real filed data can be another possible option if we can have some common field data sets.  Nevertheless, if the majority wants to study the generalization for different cell config. via simulation, we support the Ericsson’s approach to consider two different parameter sets for “small cell (UMi)” and “large cell(UMa)” environment, respectively. However, since we already have the two separate parameter sets (and also simulation results), for UMa in FR1 and for UMi in FR2, we prefer to reuse them for the generalization study on cell configuration. I.e., train the temporal domain pediction model with mixed data sets from FR1(UMa) and FR2(UMi), and evaluate the prediction accuracy of the model in each scenario. |
| Qualcomm | a), b), c), and h) | We agree with the approach proposed by Ericsson as it enables a study of model generalization across different deployment scenarios and cell sizes. |
| Xiaomi | Comments | We share similar view as Samsung. Furthermore, all these parameters have been studied by RAN1 in BM prediction. In our understanding, similar generalization performance is expected for L3 cell measurement result since the L3 cell measurement result are derived from beam measurements. We shall prioritize the mobility specific parameters for generalization study. |
| CATT | Comments | We think current effort for generalization study in the aspect of speed is enough for this study item. It seems much time consuming if we go with more genereliaztion parameters, and we have concerns whether RAN2 can finish all these simulation work on time. Also as mentioned by other companies, RAN1 has already studied similar cell parameters for generalization, and the corresponding results have been captured in TR 38.843. RAN2 should avoid duplicate work with RAN1. |
| OPPO | comments | we share the same concern with Xiaomi and CATT regarding RAN2’s work load and believe the approach proposed by Samsung is reasonable. |

### Question 2

**Question 2: any additional parameters not listed above we should consider? If you chose to suggest an additional parameter, please provide a technical justification.**

|  |  |  |
| --- | --- | --- |
| Company | Additional parameters | Comments |
| Nokia | Control of random seeds: spatial channel model, UE trajectory, etc | We have some doubts on the categorization of the above configurations into “cell parameters”. First, the generalization aspects being considered are about simulation studies, so we need to understand which part may affect the DL radio measurements for all the use cases (mainly RRM prediction). All the above listed parameters can affect path-loss, propagation. However, the discussion on control of random seeds for channel model, UE trajectory seems to be missing. We think this is also important for generalization aspects. |
| Mediatek | number of cells | Multiple cells for the data encompass all the network configuration parameters. For the generalization of data collection, it is suggested RAN2 to consider following cases:  1(Baseline): model training on data for 1 cell, model testing on data for the same cell.  2. model training on data for a set of cells, model testing on data for the same set of cells.  3. model training on data for 1 cell, model testing on data for a set of cells.  Note: the cell number can be adjusted according to company implementation. |

# 3 Proposals

[TBD]

# 5 References

[1] R2-2409652 Discussion on generalization for RRM prediction CATT, Turkcell

[2] R2-2409668 Discussion on generalization study for RRM prediction vivo

[3] R2-2409829 Discussion on Generalization Issues for AI/ML Mobility Samsung

[4] R2-2409869 Simulation Assumptions of SLS, measurement event prediction, RLF prediction and generalzatiion study MediaTek Inc.

[5] R2-2409972 Model generalization, RLF evaluation assumptions, etc. Apple

[6] R2-2410023 Discussions on evaluation methodology of AI/ML for mobility NTT DOCOMO, INC.

[7] R2-2410263 Discussion on generalization aspects Ericsson

[8] R2-2410540 Discussion on simulation assumptions and generalization Huawei, HiSilicon

[9] R2-2410800 Discussion on generalization aspects and simulation assumption ZTE Corporation

[10] R2-2410345 Discussion on other aspects of simulation assumption CMCC