**3GPP TSG RAN WG2 Meeting #131 R2-2505xxx  
Bengaluru, Indian, August 25 – 29, 2025**

**Agenda item: 8.1.2.2**

**Source: Apple, InterDigital**

**Title: Summary report of [AT131][003][AI PHY] Functionality activation (Apple)**

**WID/SID: NR\_AIML\_Air-Core– Release 19**

**Document for: Discussion and Decision**

# 1 Introduction

This is to trigger the following offline discussion:

* [AT131][003][AI PHY] Functionality activation (Apple)

Intended outcome: Agreable proposal for functionality activation open issues (interpretation, RRC-15, 17, and 45, whether for offline whether configuration is provided to L1 before model applicability, response to RAN4]]

Deadline: Thursday

# 2 Discussion

## 2.1 Agreements made on Tuesday

* RRC processing delay shouldn’t be impacted by the model loading delay
* If the UE is ready for inference by end of RRC processing delay, it reports model applicable. If not, it reports model inapplicable and doesn’t set the release flag. The network is not expected to release inference configuration (this will not be added to stage 3 specifcation).
* Once the model is applicable, UE reports applicability to network via UAI (applicable to all CSI reporting).
* Respond to RAN4

## 2.2 RAN4 LS and RRC-15

**RAN4 LS**

[R2-2505045](file:///C:\Users\wattsdy\OneDrive%20-%20InterDigital%20Communications,%20Inc\3GPP\RAN2\131%20Bangalore\Review\tdocs_131\R2-2505045.zip) LS on AI/ML functionality activation (R4-2508085; contact: CMCC) RAN4 LS in Rel-19 NR\_AIML\_air-Core To:RAN2 Cc:RAN1

*Interpretation 1 vs. 2*

[R2-2505510](file:///C:\Users\wattsdy\OneDrive%20-%20InterDigital%20Communications,%20Inc\3GPP\RAN2\131%20Bangalore\Review\tdocs_131\R2-2505510.zip) Discussion and draft reply LS on functionality activation CMCC,OPPO discussion Rel-19 NR\_AIML\_air-Core

Proposal 1: RAN2 confirm that Interpretation 1 is common understanding for AI/ML functionality activation for option A, and update the wording as follows:.

 Interpretation 1: the applicable functionalities are already considered activated (i.e., the applicable functionalities activation is completed) before upon reporting applicable functionalities via RRCReconfigurationComplete in step 4

[R2-2505192](file:///C:\Users\wattsdy\OneDrive%20-%20InterDigital%20Communications,%20Inc\3GPP\RAN2\131%20Bangalore\Review\tdocs_131\R2-2505192.zip) Discussion on LS on functionality activation and open issues on BM vivo discussion NR\_AIML\_air-Core

Proposal 1. Reply to RAN4 that interpretation 2 is aligned with RAN2 agreements. That is, UE starts to activate the applicable functionalities (i.e., the applicable functionalities activation is not completed) upon reporting applicable functionalities via RRCReconfigurationComplete.

*RRC-15: The time duration for an AI functionality to become available for inference after UE reports applicability*

[R2-2505470](file:///C:\Users\wattsdy\OneDrive%20-%20InterDigital%20Communications,%20Inc\3GPP\RAN2\131%20Bangalore\Review\tdocs_131\R2-2505470.zip) Further Discussion on the Remaining RRC Issues on LCM MediaTek Inc. discussion

Proposal 6 (RRC-17): It is clarified that loading/preparation time for the model corresponding to the applicable configuration is not considered as a component in the processing latency between the reception of RRCReconfiguration and the reporting of RRCReconfigurationComplete.

[R2-2505502](file:///C:\Users\wattsdy\OneDrive%20-%20InterDigital%20Communications,%20Inc\3GPP\RAN2\131%20Bangalore\Review\tdocs_131\R2-2505502.zip) Remaining issues on LCM procedure of UE-sided model for AI/ML based beam management Apple discussion Rel-19 NR\_AIML\_air-Core

Proposal 3 (Open issue RRC-15): On the time duration for an AI functionality to become available for inference, RAN2 conclude that it is up to UE implementation and no need of UE reporting. Whether to specify its requirement is left to RAN4.

**Rapporteur suggested way-forward:**

MediaTek P6 has been agreed on Tuesday. Thus, we can focus on P3 of Apple. It is aligned with Tuesday agreement. But it is better to further clarify that NW implicitly knows when model is ready via applicability reporting and no need of new UE reporting.

**Proposal 1 (Open issue RRC-15): On the time duration for an AI functionality to become available for inference, RAN2 conclude that it is up to UE implementation from RAN2 point of view and no further RAN2 work.**

* **Agreeable**

**Proposal 2: In Reply LS to RAN4, we can request RAN1 to check if any issue on agreement in Tuesday and related RAN2 agreements (appendix). FFS whether capture issue from RAN2 perspective.**

* **Agreeable**

**Discussion:**

**- Nokia and Ericsson think we need RAN1 to check UE beahvior agreed in Tuesday.**

**- Apple and IDC think agreemnt in Tuesday is clear and can include in LS to RAN4/RAN1**

## 2.3 RRC-17

RRC-17: Processing timing requirement of applicability/inapplicability report via RRCReconfigurationComplete

[R2-2505778](file:///C:\Users\wattsdy\OneDrive%20-%20InterDigital%20Communications,%20Inc\3GPP\RAN2\131%20Bangalore\Review\tdocs_131\R2-2505778.zip) RRC open issues for AIML for NR air interface Ericsson discussion (Moved from 8.1.1)

to determine the applicability.

Proposal 7: (RRC-17) RRCReconfigurationComplete containing applicability reports has a processing latency requirement of 16 ms with respect to the reception of RRCReconfiguration. FFS whether RAN4 input is needed. FFS whether this solves open issue RRC-15.

[R2-2505301](file:///C:\Users\wattsdy\OneDrive%20-%20InterDigital%20Communications,%20Inc\3GPP\RAN2\131%20Bangalore\Review\tdocs_131\R2-2505301.zip) Discussion on life cycle management open issues Xiaomi discussion Rel-19 NR\_AIML\_air-Core

Proposal 3: (RRC-17) Introduce multiple RRC processing delay requirements for applicability reporting based on UE capability. Values of RRC processing delay requirement is up to RAN4.

**Rapporteur suggested way-forward:**

According to Tuesday agreement “RRC processing delay shouldn’t be impacted by the model loading delay”, it precludes new RRC processing delay requirement. So, we can focus on P7 of Ericsson.

**Proposal 3: RRCReconfigurationComplete containing applicability reports has a processing latency requirement of 16 ms with respect to the reception of RRCReconfiguration, from RAN2 point of view.**

* **Almost Agreeable**

## 2.4 RRC-45

*RRC-45: How/where to capture activation of periodic inference CSI-ReportConfig in specifications*

* What to write in RRC specifications:
  1. Upon including the applicable status in RRCReconfigurationComplete, indicate to lower layers to activate the configuration, or
  2. Submit/do not submit the configuration to lower layers, which would be the equivalent of activation/no activation.

[R2-2505502](file:///C:\Users\wattsdy\OneDrive%20-%20InterDigital%20Communications,%20Inc\3GPP\RAN2\131%20Bangalore\Review\tdocs_131\R2-2505502.zip) Remaining issues on LCM procedure of UE-sided model for AI/ML based beam management Apple discussion Rel-19 NR\_AIML\_air-Core

Proposal 7 (Open issue RRC-45): As no concept of “activation/deactivation of periodic CSI reporting” in RAN1, adopt the way of “submit/do not submit the configuration to lower layers” in RRC running CR.

**Rapporteur suggested way-forward:**

P7 of Apple seems to be consensus according to company contribution. The key point is that RAN2 do not introduce new concept of “activation/deactivation of periodic CSI reporting”.

**Proposal 4: RAN2 don’t intend to introduce explicit “activation/deactivation of periodic CSI reporting”.**

* **Agreeable**
* Where exactly in RRC to capture the solution:
  1. In Section 5.3.5.3 when setting the content of RRCReconfigurationComplete, or
  2. In Section 5.3.5.3 when sending RRCReconfigurationComplete to the lower layers

[R2-2505345](file:///C:\Users\wattsdy\OneDrive%20-%20InterDigital%20Communications,%20Inc\3GPP\RAN2\131%20Bangalore\Review\tdocs_131\R2-2505345.zip) Remaining issues in LCM for BM and CSI prediction Samsung discussion Rel-19 NR\_AIML\_air-Core

Proposal 4: (RRC-45) Handling of periodic CSI-ReportConfig is described in Section 5.3.5.3 when setting the content of RRCReconfigurationComplete.

**Rapporteur suggested way-forward:**

P4 of Samsung seems to be majority view according to company contribution. In Rapporteur understanding, both a) and b) can work and their timing difference is quite minor (and it is UE’s internal processing time).

**Proposal 5: Handling of periodic CSI-ReportConfig is described when setting the content of RRCReconfigurationComplete.**

* **Agreeable**

## 2.5 (if time allows) whether configuration is provided to L1 before model applicability

If time allows:

[R2-2505838](file:///C:\Users\panidx\OneDrive%20-%20InterDigital%20Communications,%20Inc\Documents\3GPP%20RAN\TSGR2_131\Docs\R2-2505838.zip) LCM for UE-side models for beam management Ericsson discussion

Proposal 6 (RRC-45): RAN2 to ask RAN1 to resolve the decoding ambiguity at the NW regarding whether/when the UE starts sending AIML-based periodic CSI reports. Compliance with the RRC processing delay requirements should be ensured.

[R2-2505762](file:///C:\Users\panidx\OneDrive%20-%20InterDigital%20Communications,%20Inc\Documents\3GPP%20RAN\TSGR2_131\Docs\R2-2505762.zip) Remaining open issues: LCM for UE-sided model for BM use case InterDigital discussion Rel-19 NR\_AIML\_air-Core

**Proposal 2b: [RRC-45] UE does not multiplex UCI containing inference predictions until applicabile functionality is activated.**

[R2-2505502](file:///C:\Users\wattsdy\OneDrive%20-%20InterDigital%20Communications,%20Inc\3GPP\RAN2\131%20Bangalore\Review\tdocs_131\R2-2505502.zip) Remaining issues on LCM procedure of UE-sided model for AI/ML based beam management Apple discussion Rel-19 NR\_AIML\_air-Core

**Proposal 8 (Open issue RRC-45): When periodic CSI prediction is configured, it is up to NW implementation how to resolve the UCI multiplexing issue (if any).**

**Rapporteur suggested way-forward:**

If P4 of Samsung is agreeable, it means that the UE will not do periodic CSI inference until it reports RRCReconfigurationComplete. Thus, in Rapporteur understanding, NW knows when the UE will start to report CSI prediction, which may address Ericsson’s concern. We can add RAN1 action to make necessary spec changes (if any)

**Proposal X: RAN2 conclude that the UE submits the inference configuration of periodic CSI to lower layers only after the applicabilityStatus is set to applicable. Thus, NW knows that the UE will not report periodic CSI prediction until reception of RRCReconfigurationComplete. In reply LS to RAN4, add action to RAN1: “RAN1 take the RAN2 conclusion into accout and make necessary specification change (if any).”**

## 2.6 (if time allows) LS to RAN4

According to RAN2 agreement, we can reply RAN4 (cc RAN1).

## 2.7 Further discussion on when UE submits inference configuration of periodic CSI to lower layer

During F2F offline, some companies raised concern on the following proposal.

**Proposal 7 (Open issue RRC-45): As no concept of “activation/deactivation of periodic CSI reporting” in RAN1, adopt the way of “submit/do not submit the configuration to lower layers” in RRC running CR.**

The main concern is the new behaviour is not aligned with legacy UE behaviour (i.e. in legacy, upon reception of RRCReconfiguration message, the UE immediately submits RRC configuration to lower layer for validity checking and apply the configuration. Thus, above P7 was not agreeable. Thus, after official offline discussion, several interested companies continued to discuss the issue of when UE submits inference configuration of periodic CSI to lower layer. And two different options are identified:

* **Option 1: Upon reception of RRC Reconfiguration message, UE immediately submits inference configuration of periodic CSI to lower layer**
* **Option 2: Upon reception of RRC Reconfiguration message, UE holds on submitting inference configuration of periodic CSI to lower layer upon reception of RRC Reconfiguration message until applicable**

In details, we have the following detailed UE behaviour in Option 1:

|  |
| --- |
| **UE behavior of Option 1**  1> Upon receiving RRC Configuration, the configuration it is submitted to lower layers (as in legacy)  1> If Model is ready for inference within processing delay:  2> UE reports RRC configuration complete message with applicable  2> NW assumes model is available for inference and subsequent CSI reports are valid  1> else if Model is not ready within processing delay:  2> UE reports RRC configuration complete message with inapplicable with no release cause  2> If model becomes available for inference after RRC processing delay, and  2> If network has not released configuration:  3> UE reports updated applicability via UAI (according to RAN2 agreements)  3> Network sends RRC reconfiguration message again (to support RAN1 specification)  3> UE sends RRC reconfiguration complete message indicating applicability (according to RAN1 specification) |

And detailed UE behaviour in Option 2 is:

|  |
| --- |
| **UE behavior of Option 2**  1> Upon receiving RRC Configuration, the configuration it is not submitted to lower layers (change in legacy)  1> If Model is ready for inference within processing delay:  2> UE reports RRC configuration complete message with applicable  2> UE submits CSI config down to lower layers  2> NW assumes model is available for inference and subsequent CSI reports are valid  1> else if Model is not ready within processing delay:  2> UE reports RRC configuration complete message with inapplicable with no release cause  2> If model becomes available for inference after RRC processing delay, and  2> If network has not released configuration:  3> UE reports updated applicability via UAI (according to RAN2 agreements)  3> Network sends RRC reconfiguration message again (to support RAN1 specification)  3> UE submits CSI config down to lower layers  3>  UE sends RRC reconfiguration complete message indicating applicability (according to RAN1 specification) |

Rapporteur think both solutions can work, but they have pros and cons, as summarized in Table.1.

|  |  |  |
| --- | --- | --- |
|  | Option 1 | Option 2 |
| Pros | Same as legacy UE behaviour |  |
| Cons | May lead to one additional ambiguity time (If the model is not ready for inference after RRC processing delay and the network does not release the configuration, this ambiguity continues until a subsequent report is sent) | 1. Derive from legacy UE behaviour  2. Until the NW receives successfully the applicability reports in RRCReconfigurationComplete, the NW cannot know whether the UE started sending CSI reports or not. |
| RAN2 impact | N/A  (if RAN2 don’t pursue enhancement that UAI with applicability can allow UE autonomously to start inference) | New UE behavior upon receiving RRC Reconfiguration message: the RRC may need to parse the CSI-MeaseConfig to maintain the configurations relevant to AIML to determine applicability. |
| Potential RAN1 impact | Handle UCI multiplexing issue during additional ambiguity time | N/A |
| Potential RAN3 impact | Exchange inapplicability indication from the CU to the DU | Exchange applicability/inapplicability indication from the CU to the DU |

**Table.1 Analysis between Option 1 and Option 2**

We need to have way-forward because CB is in this afternoon. And reviewing company comments, it seems there are still different understanding on the two options and it may be difficulty to down-select. To make progress, Rapporteur would like to try whether company can accept the following way-forward:

**Proposal 6: RAN2 ask RAN1 in LS: On the following RAN1 note agreed in RAN1:**

***Note: UE is not expected to be configured with a CSI-ReportConfig for inference configuration for a non-applicable set of inference parameters or a non-applicable CSI-ReportConfig***

**What is RAN1’s understanding on how UE handles CSI-ReportConfig when CSI-ReportConfig for inference configuration is not applicable.**

**Proposal 7: RAN2 confirm that after sending UAI when becoming applicable from inapplicable, the UE needs to wait NW to send a followed RRCReconfiguration with inference configuration.**

**Proposal 8: Postpone RAN3 impact discussion. Interested company can raise the issue in RAN3 directly.**

# 3 Conclusion

**Need further discussion**

**Proposal 6: RAN2 ask RAN1 in LS: On the following RAN1 note agreed in RAN1:**

***Note: UE is not expected to be configured with a CSI-ReportConfig for inference configuration for a non-applicable set of inference parameters or a non-applicable CSI-ReportConfig***

**What is RAN1’s understanding on how UE handles CSI-ReportConfig when CSI-ReportConfig for inference configuration is not applicable.**

**Proposal 7: RAN2 confirm that after sending UAI when becoming applicable from inapplicable, the UE needs to wait NW to send a followed RRCReconfiguration with inference configuration.**

**Proposal 8: Postpone RAN3 impact discussion. Interested company can raise the issue in RAN3 directly.**

**Agreeable**

**Proposal 1: On the time duration for an AI functionality to become available for inference, RAN2 conclude that it is up to UE implementation from RAN2 point of view and no further RAN2 work.**

**Proposal 2: In Reply LS to RAN4, we can request RAN1 to check if any issue on agreement in Tuesday and related RAN2 agreements (appendix). FFS whether capture issue from RAN2 perspective.**

**Proposal 4: RAN2 don’t intend to introduce explicit “activation/deactivation of periodic CSI reporting”.**

**Proposal 5: Handling of periodic CSI-ReportConfig is described when setting the content of RRCReconfigurationComplete.**

**Proposal 3: RRCReconfigurationComplete containing applicability reports has a processing latency requirement of 16 ms with respect to the reception of RRCReconfiguration, from RAN2 point of view.**

# 4 References

[1] RP-243244, Revised WID: Artificial Intelligence (AI)/Machine Learning (ML) for NR Air Interface, Qualcomm.

[2] RAN2#130, Chair Notes.

[3] R2-2505778, Remaining RRC open issues in feature AIML PHY, Ericsson.

# Appendix

The UE performance requirements for RRC procedures are specified in the following tables. The performance requirement is expressed as the time in [ms] from the end of reception of the network -> UE message on the UE physical layer up to when the UE shall be ready for the reception of uplink grant for the UE -> network response message with no access delay other than the TTI-alignment (e.g. excluding delays caused by scheduling, the random access procedure or physical layer synchronisation). In case the RRC procedure triggers BWP switching, the RRC procedure delay is the value defined in the following table plus the BWP switching delay defined in TS 38.133 [14], clause 8.6.3.

A red arrow pointing to a line

AI-generated content may be incorrect.

Figure 12.1-1: Illustration of RRC procedure delay