Qualcomm

RP-233131

Views on scope for AI/ML for NR Air Interface in Rel-19

AI: 9.1.1.1

3GPP TSG RAN meeting #102

Edinburgh, Scotland

December 11-15, 2023

Rel-19

Proposed high-level scope

- Normative phase of recommended 1-sided cases (BM+POS) – down-select sub use cases
  - LCM Management [RAN2/RAN1]
  - Optimizations for specific use cases [RAN1]
  - Core requirements [RAN4]
  - Performance requirements (including test cases) [RAN4]
- RAN4 to continue study on testability (1-sided & 2-sided models)
  - Checkpoint in Sept ‘24
- RAN1 to continue study on 2-sided compression and CSI-prediction
  - Address open issues identified for CSI compression
  - Address open issues identified for CSI prediction
  - Checkpoint in Sept ‘24 on how to proceed
- RAN2 to reply to SA2's LS
- TR updates based on studies

Beam management

Rel-19 normative work

- **Proposal**: Rel-19 normative work on the following
  - Both **BM-Case1** and **BM-Case2**:
    - **BM-Case1**: Spatial-domain DL Tx beam prediction for Set A of beams based on measurement results of Set B of beams
    - **BM-Case2**: Temporal DL Tx beam prediction for Set A of beams based on the historic measurement results of Set B of beams
  - DL Tx beam prediction for both **UE-sided model** and **NW-sided model**
  - Necessary **signaling/mechanism**(s) to facilitate **data collection**, **model inference**, and **performance monitoring** for both UE-sided model and NW-sided model, including
    - Enhanced L1 report enabled by UE-side AI/ML inference
    - RS configuration to facilitate UE side data collection, model inference, and performance monitoring
    - Assistance information from NW to UE to help UE categorize data
  - **Signaling/mechanism**(s) to facilitate necessary **LCM operations** via 3GPP signaling for UE-sided model, including
    - Methods to ensure consistency across training and inference, including consistency in association/mapping of Set A beams and Set B beams for UE-side models, considering NW-side additional conditions

Positioning accuracy enhancements

Rel-19 normative work

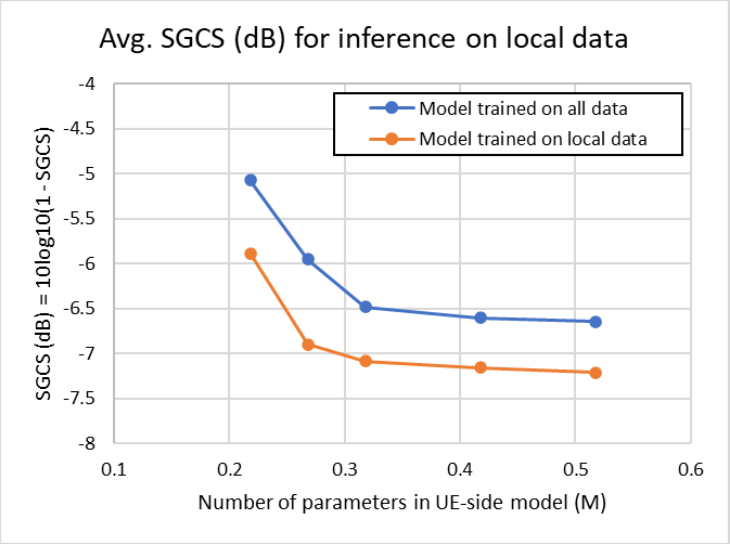
- **Proposal**: Specify **Direct AI/ML Positioning** and **AI/ML Assisted Positioning** in R19 WI:
  - Direct AI/ML Positioning (at least Case1) and AI/ML Assisted Positioning (Case2a and Case3a)
  - Specify necessary **signalling** of **data collection** and **related other information**:
    - e.g., assistance information from LMF to UE to help UE categorize data and generate label; signaling for RS configurations; data collection quality requirements, etc.
  - Specify necessary **signalling** and details for **measurement enhancements**:
    - e.g., Prioritize reusing existing measurement; considering tradeoff of signaling overhead & achieved positioning accuracy
  - Specify necessary **signalling** details of **monitoring method**(s):
    - e.g., labeling assistance from LMF; configuring model failure detection and related signaling
  - Specify necessary **signalling** details and applicability of **generalization method**(s):
    - e.g., assistance signaling from LMF to UE to help UE: identify positioning scenario/ (mixed) dataset, select/switch model, apply finetuning and labeling; considering applicability/feasibility of generalization to positioning with non-ideal settings (e.g., non-ideal labeling in mixed dataset & finetuning, non-ideal scenario identification & model selection, non-ideal model switching, etc.)

CSI compression

Rel-19 RAN1 to continue study and revisit at checkpoint

- **Address** the following aspects that led to the **lack of RAN1 consensus** on the recommendation for normative work:
  - Trade-off between performance and complexity/overhead.
  - Issues related to inter-vendor training collaboration.
- Recommended directions (study objective):
  - Temporal domain aspects
    - e.g., temporal compression, joint prediction and compression, use of past CSI-RS, differential feedback
  - Localized models
  - Complexity/performance tradeoff
  - Specification and collaboration flavors that reduce inter-vendor logistical complexity

Localized models can achieve the same performance with significantly lower complexity



Source: R1-2305328, RAN1#113

CSI prediction

Rel-19 RAN1 to continue study and revisit at checkpoint

- **Address** the following aspect that led to the lack of RAN1 consensus on the recommendation for normative work:
  - Lack of results on the performance gain over non-AI/ML based approach and associated complexity.
- Recommended directions (Study objective):
  - Further evaluation of AI/ML relative to a non-AI/ML based approach
  - Further evaluation covering aspects such as generalization across scenarios, scalability across configurations
  - Rel-18 MIMO temporal domain CSI feedback as a baseline

Data Collection

Continue study & possible architecture impact

- RAN2 discussions on UE data collection concluded by capturing that:
  - *The following proposals were discussed in RAN2*
    1. *UE collects and directly transfers training data to the OTT server*
       1a) OTT (3GPP transparent)
       1b) OTT (non-3GPP transparent)
    2. *UE collects training data and transfers it to CN. CN transfers the training data to the OTT server.*
    3. *UE collects training data and transfers it to OAM. OAM transfers the needed data to the OTT server.*
  - *RAN2 did not study or analyze the proposals and did not agree to requirements or recommendations.*
- Since much of this discussion is outside RAN expertise, since SA2 has offered to help in LS (S2-2311921/RP-232727), and that we have concerns with options 2 and 3, we propose that:
  - RAN reply to SA2 requesting help in designing:
    - A secure UE data collection between the UE and an OTT server (option 1 above)
    - A secure UE model transfer between the OTT server and the UE
    - Model ID management that ensures uniqueness

Testability and Interoperability

RAN5 involvement

- AI/ML testability/interoperability is raising new challenges in RAN2/4 & RAN5
  - New test setup/procedures will likely be needed for all use cases under discussion.
  - Testing options identified for two-sided model could lead to a change in 3GPP testing framework/paradigm.
- Close RAN2/4/5 interaction in the discussion might be needed to enable progress and identify the best solutions
  - RAN5 has no precedence of formally participating in SI/WI led by other workgroups.
  - RAN5 possible participation framework
    - RAN2/4 identifies the list of items that need RAN5 WG study.
    - RAN5 is formally added to the objectives of the SI/WI.

Qualcomm

Thank you!

Follow us on:

For more information, visit us at:

www.qualcomm.com & www.qualcomm.com/blog

Nothing in these materials is an offer to sell any of the components or devices referenced herein.

©2018 Qualcomm Technologies, Inc. and/or its affiliated companies. All Rights Reserved.

Qualcomm is a trademark of Qualcomm Incorporated, registered in the United States and other countries. Other products and brand names may be trademarks or registered trademarks of their respective owners.

References in this presentation to “Qualcomm” may mean Qualcomm Incorporated, Qualcomm Technologies, Inc., and/or other subsidiaries or business units within the Qualcomm corporate structure, as applicable. Qualcomm Incorporated includes Qualcomm’s licensing business, QTL, and the vast majority of its patent portfolio. Qualcomm Technologies, Inc., a wholly-owned subsidiary of Qualcomm Incorporated, operates, along with its subsidiaries, substantially all of Qualcomm’s engineering, research and development functions, and substantially all of its product and services businesses, including its semiconductor business, QCT.