R4-2400504

3GPP TSG-RAN WG4 Meeting #110
Athens, Greece, 26 Feb - 01 March, 2024
Agenda item: 14
Source: Apple
Release: Rel-19
Document for: Discussion

# AI/ML For RRM Measurement Enhancement Rel-19

# Rel-19 AI/ML based RRM enhancement: L3 Measurement Enhancement

- Motivation:
  - **Delay Reduction for FR2**: The L3 measurement delay depends on the number of measurement samples (to reduce channel variations), the Rx beam sweeping factor and the SMTC periodicity.
    - For example, the delay is computed as: 3( # number of measurement samples) x 8 (#Rx beams) x $T_{SSB}$ = 24 $T_{SSB}$. AI/ML can be used to reduce the measurement delay ( $T_{SSB}$ is the SMTC periodicity e.g. 20ms)
    - Delay reduction could be achieved by estimating RSRPs across subsampled Rx beams by spatial estimation or estimating RSRPs across subsampled measurement periods by temporal estimation
  - **Reduce the measurement overhead for FR2**: The overhead of RS used for RRM (i.e., SSB and CSI-RS) is quite high in the current NR design, e.g., when SMTC periodicity is 20ms, the related SMTC overhead in FR2 is 25%, especially for FR2 (with 5ms SSB burst). Reducing the measurement overhead will lead to:
    - *Reduced scheduling restrictions*.
    - *Reduced measurement restrictions*
  - **Reduce complexity of measurements**:
    - By skipping measurements for multiple TX beams (SSBs) the complexity of the RRM measurements will be decreased.
    - By estimating RSRPs across cells and across frequencies from other cell/frequency measurements

# AI/ML based RRM Enhancement: RAN4's Distinctive Role

- Differences between RAN1 R18 study, RAN2 R19 study and RAN4
  - RAN4 scope for estimating L3-RSRPs differ from RAN1's scope for BM in the following aspects:
    - ***Low operating SNR***: L3 RSRP measurements apply to both **serving and neighboring cells**, including intra-/inter-frequency. RRM Measurements at neighbor cells impose very low SNR conditions
    - ***Model generalization***: Different deployment scenarios (NW vendor/cell-site/ frequency band) would have their own spatial/frequency/temporal characteristics. Assistance information can aid the generalization of the same AI/ML model across different deployments. In other cases where generalization cannot be guaranteed a different AI/ML model will be trained and a new AI/ML model ID will be assigned
    - ***UE oriented enhancement***: RAN1 focuses on predicting RSRPs for selecting the best TX beams. RAN4 focuses on RX beam reduction and measurement gap reductions which lead to delay reductions
  - RAN4 scope for estimating L3-RSRPs differ from RAN2's proposed scope for mobility
    - RAN2 R19 study focus on RRM/RSRP prediction
    - RAN4 focus on **reducing the overhead/delay/complexity via enhanced RSRP estimation.**
- Proposed objectives of the study
  - **L3-RSRP Enhancement (estimating RSRP in spatial and temporal domain for selecting optimal TX/RX beam pairs)**
  - **L1-RSRP LTM (Layer 1/2 triggered mobility) Enhancement (estimating RSRP in spatial and temporal domain for selecting optimal RX beam pairs)**

# Rel-19 AI/ML Based RRM Enhancement: Challenges

- ***SNR/Distortion (perturbations for the same deployment)***
  - Training dataset consists of low resolution images (RSRP map) as “before” images and the corresponding high resolution images as “after” images. “Before” images include different degradation types, namely additive white Gaussian noise (AWGN), antenna spacing mismatches, beam width mismatches between trained data and real data  etc
  - Although learning-based restoration processors could have high performance, they are very sensitive to the perturbation of the degradation model. The AI/ML model can easily fail to restore the RSRP image whose degradation model is slightly different from the trained degradation mode. It would be challenging/unrealistic to train with with all kind of degradation types/levels.
- ***Generalization of the AI/ML model to different deployment***
  - *Diverse deployment*: AI/ML Spatial estimation should be employed across multiple deployment scenarios. Different cells (Serving or Neighbor) could have different Tx beamforming implementations, different codebook sizes, and/or the order of Tx beam sweeping could differ, different power delay profiles.  Also, AI/ML model should also be employed for measurements across inter-frequency neighbor cells
    - *Assistance information:*
      - The number of AI/ML models should be kept low in order to reduce the complexity/ model storage and AI/ML model transfer requirements
      - In order to enhance the generalization capacity of a specific AI/ML model, assistance information could be used to support AI/ML functionality across different deployment scenarios (infra vendor/ cell site/ frequency )
    - *AI/ML Model ID*
      - In case the AI/ML model’s generalization cannot be guaranteed across a different deployment scenario, a new offline training procedure could take place and the new model shall be transferred to UE with a new AI/ML model ID assigned

# Illustration of FR2 Delay Reduction Through Spatial/Temporal Estimation

# Illustration Rel-19 AI/ML Based RRM: Inter Frequency RRM measurement estimation

- Use AI/ML to predict RRM measurements across different frequencies
- Measure intra-frequency Beams from Cell A: A{1,3,7} and from Cell C: C{3,5,6,8}
- AI/ML uses spatial/frequency measurements to estimate inter-frequency beams for cell B
- Inter frequency RRM measurement estimation will lead to RRM measurement overhead reduction and alleviate the need for measurement gaps

Spatial/Frequency Domain Estimation

# Simulation Results: AI/ML Spatial Estimation at High SNR

- The labels $aTx/bRx/c\,\%\,TR/N = d/CH$ in the figure denote:
- $a$ is the number of subsampled Tx beams
- $b$ is the number of subsampled Rx beams
- $c = (1 - \frac{ab}{N_{Rx}M_{Tx}})100\,\%$ is the TR
- N=$d$ is the number of top-N beams
- $CH = \{LOS, NLOS\}$ is the channel type

TR: Training Reduction

Network is trained with antenna spacing = $\lambda/2$

Significant overhead reduction can be achieved with AI/ML estimation
90% measurement reduction with accuracy > 85%

# Low SNR: Beam Accuracy as a function of Top-K beam and SNR levels

- For low SNR, the value of K in the top-K reported beams could be high. The importance of selecting a higher number for the top-K beam selection is supported below by simulations results.
- In order to achieve >80% beam accuracy we need top-K >16 beams for SNR=-5dB.
- Increasing the number of top-K will lead to significant feedback overhead, thus posing a challenge for low SNR

TM and © 2024 Apple Inc. All rights reserved.