vivo

3GPP TSG RAN Meeting #101 RP-231789

Bangalore, September 11 - 15, 2023

Agenda Item: 8A.2.1

Source: vivo

Title: Rel-19 items for AI/ML over Air interface

Document for: Discussion and decision

Guidance from RANP #100

AI/ML Air Interface

- Whether/how to continue Artificial Intelligence (AI)/Machine Learning (ML) air interface in Rel-19 depends on the outcome of the Rel-18 study.
- Normative work on AI/ML air interface is expected upon successful completion of the Rel-18 study
  - In particular, which are the set of sub use cases for the potential WI?
- Whether or not to have parallel study, e.g., new use case(s) (e.g., mobility led by RAN2?)
- Important to keep the total amount of work related to AI/ML air interface within a reasonable load in Rel-19

Topics

- Rel-19 WI or SI for Rel-18 continuation?
- Rel-19 scope for Rel-18 continuation
- New use cases for Rel-19 AI/ML over air interface
- Coordination with SA

Rel-19 air interface AI/ML WI or SI for Rel-18 continuation?

- Based on RAN #100 guidance, Rel-19 WI is expected as long as Rel-18 SI is successfully completed.
- Completion level of Rel-18 SI till RAN #101 is expected at least 80% with the following remaining issues to be finished within Q4 for RAN1/RAN2/RAN4
  - RAN1 has already completed the following agendas
    - Beam prediction evaluations and specification impact analysis
    - Positioning evaluation and specification impact analysis
    - CSI enhancement performance evaluation
  - The following aspects are expect to be extended to Q4
    - Framework: functionality/model identification, model transfer pros/cons analysis, data collection details;
    - CSI enhancement: training collaboration comparison, model identification related procedures;
  - RAN2 remaining issues include
    - Further analysis of solutions for model transfer/delivery
    - Down-selection of Data collection framework based on RAN1 input
    - Converge on functionality mapping based on RAN1 input
    - Procedures of model/functionality identification based on RAN1 conclusion
  - RAN4 remaining issue include
    - Analysis on the different options for 2-sided model testing options
    - Analysis on reference model for defining requirements for both 2-sided and 1-sided model
    - KPI/metrics and testing procedures for use cases;
    - Testing options for verifying generalization performance
- Rel-18 SI is expected to be successfully concluded in Q4 by completing above issues.
- **Proposal: A Rel-19 air interface WI is expected from start of Rel-19 for Rel-18 SI continuation.**

Rel-19 scope for Rel-18 continuation

- Based on RAN #100 guidance, Rel-19 scope depends on Rel-18 outcome:
  - Rel-19 AI/ML is one of the most important features in 5.5 G and beyond.
  - It is expected that those techniques identified as beneficial in Rel-18 study should be explored as much as possible in Rel-19
- Those techniques that are identified being beneficial/necessary or having gains can be included within Rel-19 scope.
  - This is applicable for both framework discussion and use case discussion.

Rel-19 scope for Rel-18 continuation

Framework and LCM components

- In Rel-18 study, RAN1/RAN2/RAN4 identifies the following components for LCM and focus the study in part of the following:
  - Data collection
    - Note: This also includes associated assistance information, if applicable.
  - Model training
  - Functionality/Model identification
  - Model transfer
  - Model inference operation
  - Functionality/Model selection, activation, deactivation, switching, and fallback operation.
  - Functionality/Model monitoring
  - Model update
  - UE capability
- Above highlighted part are either necessary or identified as beneficial for AI/ML framework
- **Proposal: Support to include the following LCM components for framework design for Rel-19 AI/ML WI**
  - **Data collection, Functionality/Model identification, Model transfer, Model inference operation, Functionality/Model monitoring, UE capability, Functionality/Model selection, activation, deactivation, switching and fall back operation,**

Rel-19 scope for Rel-18 continuation

Framework

- Rel-19 AI/ML work for air interface is paving way for 6G, the design of framework should be inclusive covering all major aspects well studied in Rel-18
  - Considering both the cases data collection is done at UE side and at the network side
    - For UE sided model, data collection is at UE side and/or network side;
    - For network sided model, data collection is at network side with assistance from UE;
    - For two sided model, data collection is at network side and/or UE side;
  - Considering both the cases model storage entities are at UE side or network side;
    - For UE sided model, model storage is at the network side and/or at UE side;
    - For network sided model, model storage is at the network side;
    - For two sided model, model storage is at the network side and/or UE side;
  - Considering both collaboration level y and level z
    - For the case that model is stored at the network, collaboration level z is envisioned;
- **Proposal: Based on Rel-18 study and targeting a future-proof framework, Rel-19 WI for Rel-18 continuation should assume that data collection can be done at both UE side and network side, model storage can be done at UE side and network side and both collaboration level y and level z are considered.**

Rel-19 scope for Rel-18 continuation

Use cases

- The following use cases are studies in Rel-18:
  - CSI: CSI compression (mean UPT up to 15% for max rank 2, 5% UPT up to: 20.9% for max rank 1, 14.9% for max rank 2), CSI prediction
  - Beam prediction: spatial domain, temporal domain; (3/4 overhead reduction with marginal performance loss)
  - Positioning: Assisted, Direct (positioning accuracy of <1m at CDF=90% compared to >10m for legacy)
- All use cases show gains compared with legacy:
  - CSI compression is with less gain (10~20%) compared to other cases but can be easily extended to time domain compression with larger gains; (More results can be found in Appendix1)
  - CSI compression is the only two sided case among the six sub use cases;
- **Proposal: Consider all six sub use cases studied in Rel-18 to be included in Rel-19 WI;**
- **Proposal: 4 TUs are expected for Rel-19 WI RAN1 part to cover all six sub use cases;**

Rel-19 scope for Rel-18 continuation

RAN2 work analysis

- RAN2 is expected to further analysis and down-select the solutions in Rel-19.
  - Data collection frameworks for different AI purposes of each use case
  - Model transfer/delivery for different deployment scenarios
    - Specify model transfer from NW (e.g., gNB/LMF) to UE, CP- and/or UP-based solutions can be considered
  - Model/functionality identification
    - Taking capability exchange as starting point, additional enhancement is expected to support flexible model info update
    - Coordinate with SA WGs is needed on the definition and allocation of unique model ID
- Besides, RAN2 will specify the signaling/procedure according to RAN1 conclusion:
  - Functionality/Model selection, activation, deactivation, switching, and fallback operation
  - Use case specific signaling, e.g., RS configuration.

- **Proposal: 2 TUs are expected for Rel-19 WI RAN2 part to specify the essential LCM procedures;**

New use cases for Rel-19 AI/ML for air interface

Overview

- Rel-18 studied PHY layer use cases. Rel-19 should consider more use cases that can further expand the application of AI/ML in air interface :
  - Higher layer use cases
  - Use cases that utilize features not directly related to channel
    - Rel-18 PHY use cases mainly explore features related to channel propagation conditions;
- Candidate use cases include:
  - AI based mobility enhancements (RAN2-led)
  - PA efficiency/nonlinearity handling improvement, including e.g., one-sided or two-sided operation (RAN4-led)

New use cases for Rel-19 AI/ML for air interface:
AI based mobility enhancement

- AI based mobility enhancements (RAN2-led)
  - AI based handover optimization
    - Handover delay caused by legacy reporting criteria (e.g., TTT) may lead to poor user experience at source cell, or even RLF/HOF due to too late handover
    - Legacy solutions to reduce handover delay by reducing TTT duration or using R18 L1/L2-triggered mobility may result in other unintended events, e.g., too-early HO, ping-pong HO, especially for the high-speed UEs
    - AI based mobility enhancement can reduce handover delay and rate of unintended events (please check Appendix 2 for more evaluation and field test results)
  - AI assisted RRM measurement
    - Lots of beams/cells need to be measured in intra/inter frequency scenario for mobility, which brings increaded measurement effort and power consumption
    - AI assisted RRM measurement can reduce the measurement effort and power consumption (please check Appendix 2 for more evaluation and field test results)
- **Proposal: Support a RAN2-led Study Item for AI based mobility enhancement in Rel-19.**

New use cases for Rel-19 AI/ML for air interface:

AI based mobility enhancement

- The following work is expected for the study:
  - Use case discussion:
    - At least RRM measurement (e.g., RSRP, SINR) prediction, target Cell prediction, unintended events prediction should be considered
    - Both L3 mobility and L1/L2 triggered mobility should be considered
    - Representative sub-use cases need to be selected for further evaluation and specification impact analysis based on expected performance gains and necessity of specification work
  - Performance evaluation of selected representative use cases:
    - For the selected representative sub-use cases, RAN2 needs to take detailed evaluations to achieve common understanding and observation on gains and target scenarios/configurations;
  - Potential specification impact analysis
    - Delta to the framework studied in Rel-18 should be considered;
- **Proposal: At least 2TU is needed to finish the related study for AI based mobility enhancement in Rel-19**.

New use cases for Rel-19 AI/ML for air interface:

Potential SID objective for AI based mobility enhancement

- **Proposal: A RAN2-led 2TU SI is expected with following objectives for AI based mobility enhancement:**
  - Study and finalize the sub use cases from the following for evaluation and specification impact analysis for AI/ML based mobility optimization: [RAN2]
    - RRM measurement (e.g., RSRP, SINR) prediction, target Cell prediction, unintended events prediction
    - UE sided model [network sided model and two sided model]
  - For the selected sub-use case for evaluation, evaluate performance benefits of AI/ML based algorithms [RAN2]
    - Methodology based on statistical models (from [TR 38.901]), for system level simulations.
    - KPIs: Determine the common KPIs and corresponding requirements for the AI/ML operations.
  - Assess potential specification impact, specifically for the selected sub-use cases, including [RAN2, RAN1, RAN4]
    - Identify AI/ML framework applicable for the selected sub-use case
      - Identify applicable levels of collaboration between UE and NW
      - Characterize lifecycle management of AI/ML based mobility optimization: e.g., model training, model deployment, model inference, model monitoring, model updating
      - Data collection aspects
      - Note: Federal learning can be considered for model training.
    - PHY layer aspects and protocol aspects of the identified framework
      - E.g, identify impacts of different collaboration levels, the input and output for model training and model inference purpose, the performance metrics for model monitoring purpose
    - Interoperability and testability aspects

New use cases for Rel-19 AI/ML for air interface:

AI based PA efficiency/nonlinearity handling improvement

- AI based PA efficiency/nonlinearity handling improvement (RAN4-led)
  - PA non-linearity has large performance impact to UL transmission;
    - Higher power back-off decreasing transmission power and reduces UL coverage/throughput
    - PA non-linearity impairments becomes heavier in larger band width
  - Legacy methods to handle PA non-linearity includes
    - PAPR reduction techniques, e.g., tone reservation (generating peak canceling signal for the reserved tones), or SLM (select the scrambled signal from all scrambled signal set with lowest PAPR), or
    - Digital predistortion, e.g., DPD based on look up table or formulated on Volterra series with estimated parameters.
  - AI based PA efficiency/nonlinearity handling can further improve the performance for PAPR reduction and digital distortion compared to legacy (Appendix 3)
- **Proposal: Support a RAN4-led Rel-19 study item for AI based PA efficiency/nonlinearity handling improvement;**

Coordination with SA

- At least the following areas are identified with SA impact for the three studied use cases:
  - AI + Positioning procedures and signaling
  - Model identification procedures and signaling
  - Model transfer/Data collection related issues;
- Lessons learned are that RAN should coordinate with SA as early as possible:
  - Rel-18 RAN study item did not trigger any SA2 work successfully and SA2 were not involved in Rel-18 AIML study phase
  - SA2 have to start the discussion from scratch and SA2 study work expected would be large in R19 normative phase.
- **Proposal: In R19 RAN should coordinate with SA to study and specify the corresponding solutions for application of AI/ML in air interface.**

Summary of views for Rel-19 AI/ML work

- **A Rel-19 air interface WI is expected from start of Rel-19 for Rel-18 SI continuation with the following work**
  - Consider techniques identified as beneficial and necessary from Rel-18 study;
  - Support to include the following LCM components for framework design for Rel-19 AI/ML WI
    - Data collection, Functionality/Model identification, Model transfer, Model inference operation, Functionality/Model monitoring, UE capability, Functionality/Model selection, activation, deactivation, switching and fall back operation,
  - Based on Rel-18 study and targeting a future-proof framework, Rel-19 WI for Rel-18 continuation should assume that data collection can be done at both UE side and network side, model storage can be done at UE side and network side and both collaboration level y and level z are considered
  - Consider all 6 sub-use cases;
  - RAN1 needs 4TU for work identified； RAN2 needs 2TU for work identified
- **A RAN2 led SI for AI/ML based mobility enhancement is expected in parallel**
  - Evaluation and specification impact analysis are both included;
  - 2TUs are expected for corresponding work;
- **Support a RAN4-led Rel-19 study item for AI based PA efficiency/nonlinearity handling improvement;**
- **In R19 RAN should coordinate with SA to study and specify the corresponding solutions for application of AI/ML in air interface.**

vivo

THANK YOU.
谢谢。

Appendix1: Temporal-spatial-frequency domain CSI compression

- Principle: exploiting temporal correlation in consecutive CSIs to further improve the performance gains

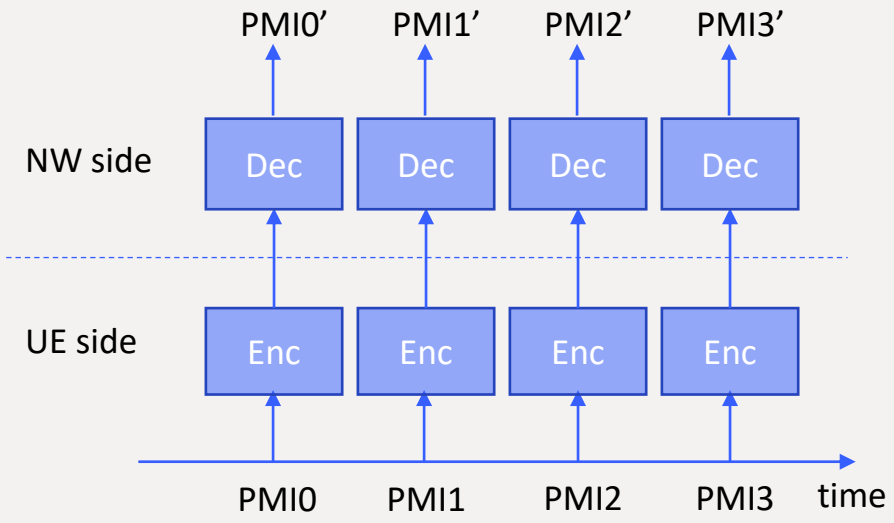


Illustration of spatial-frequency domain CSI compression in R18

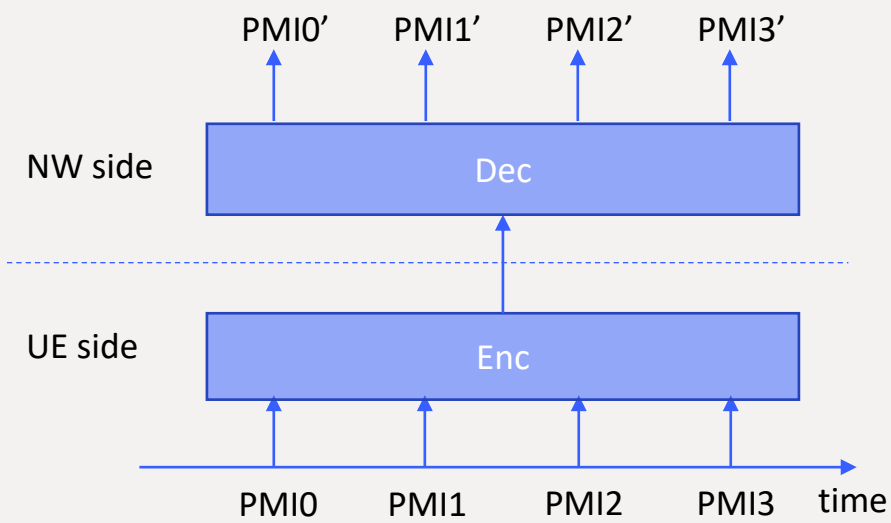


Illustration of temporal-spatial-frequency domain CSI compression

Appendix1: Temporal-spatial-frequency domain CSI compression

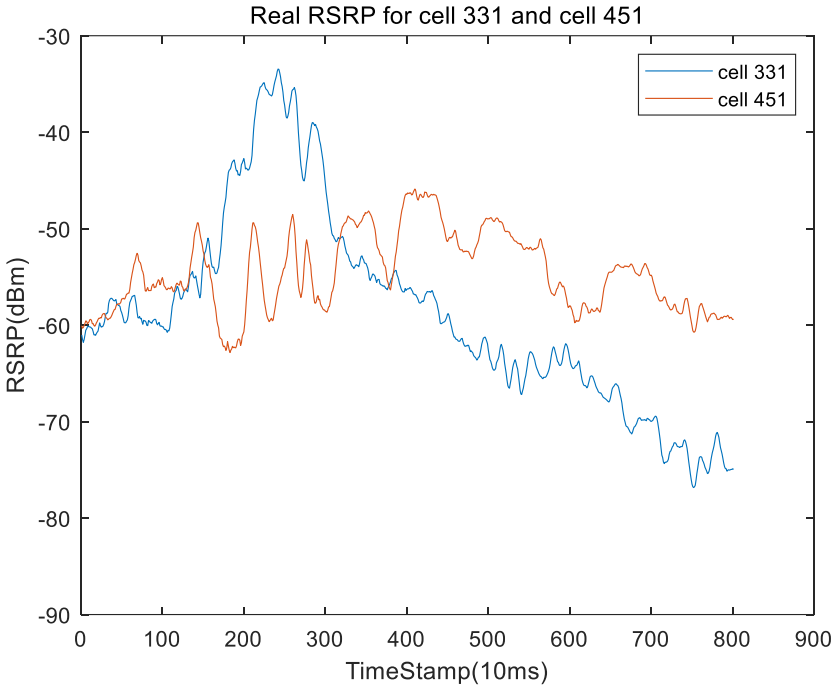
- Initial results on joint compressing 4 PMIs (rank1 considered)

	SGCS on PMI0	SGCS on PMI1	SGCS on PMI2	SGCS on PMI3	Averaged SGCS
Legacy R16 codebook	0.8172	0.8170	0.8171	0.8172	0.8172
Benchmark: s-f model with payload 64*4	0.9071 (+11.02%)	0.9070 (+11.01%)	0.9074 (+11.05%)	0.9072 (+11.01%)	0.9072 (+11.01%)
Joint compression model with payload 64*4	0.9522 (+16.51%)	0.9610 (+17.62%)	0.9613 (+17.64%)	0.9573 (+17.14%)	0.9579 (+17.21%)
Joint compression model with payload 64*2	0.9250 (+13.21%)	0.9368 (+14.66%)	0.9393 (+14.95%)	0.9321 (+14.06%)	0.9351 (+14.42%)
Joint compression model with payload 64	0.8747 (+7.04%)	0.8933 (+9.33%)	0.8937 (+9.37%)	0.8796 (+7.63%)	0.8853 (+8.33%)

- Exploiting temporal domain correlation further provides ~6% relative SGCS gain on top of s-f domain compression, and the total improvement over legacy R16 CB is ~17% SGCS (~50% to ~60% additional gain)
- Overhead further reduces ~60% compared with s-f model, and the overall overhead reduction compared with R16 CB improves from ~30% to ~70%

Simulation parameters: Uma 38.901, 100 users with spatial consistency, carrier frequency 3GHz, subcarrier spacing 30KHz, 13 subbands (10MHz, 4RBs/subband), 32 gNB antenna ([Mg Ng M N P; Mp Np] = [1 1 2 8 2; 2 8]), 2 UE antenna ([Mg Ng M N P; Mp Np] = [1 1 1 1 2; 1 1]), 100% outdoor UE, 30km/h, layer0 considered.

Appendix2: AI+Mobility field test

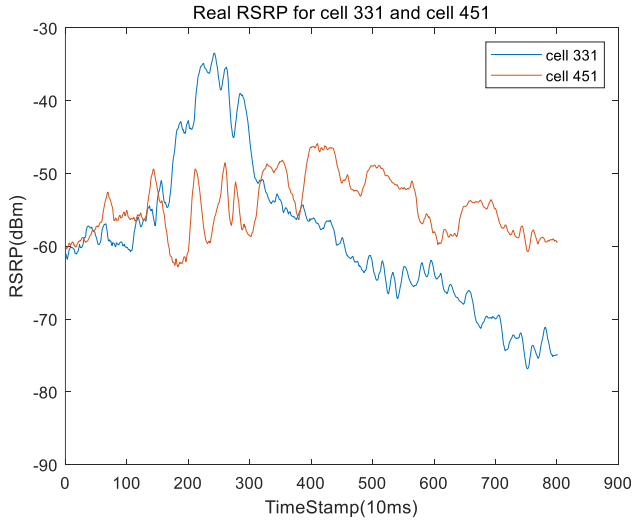


Carrier Frequency: FR1, 3.5GHz UE speed: 80~120km/h

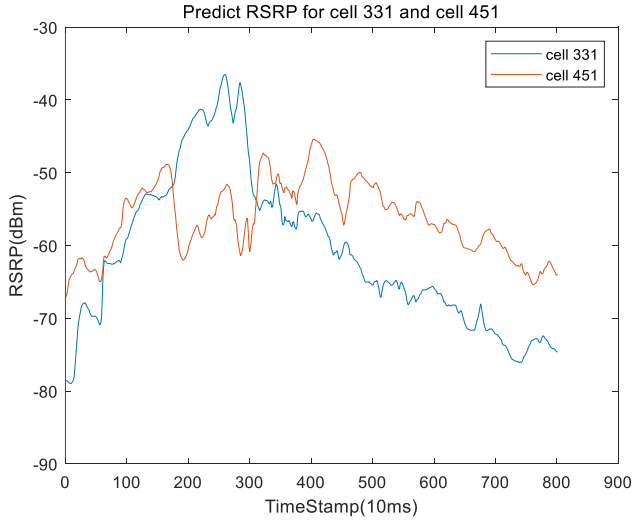
- During UE high-speed movement, the RSRP of the neighbor may becomes better than the serving cell in a short period of time.
- UE will handover to the neighbor cell and handover back quickly to the last serving cell, i.e., ping-pong handover occurs.

Appendix2: AI+Mobility field test

- **Field test RSRP**

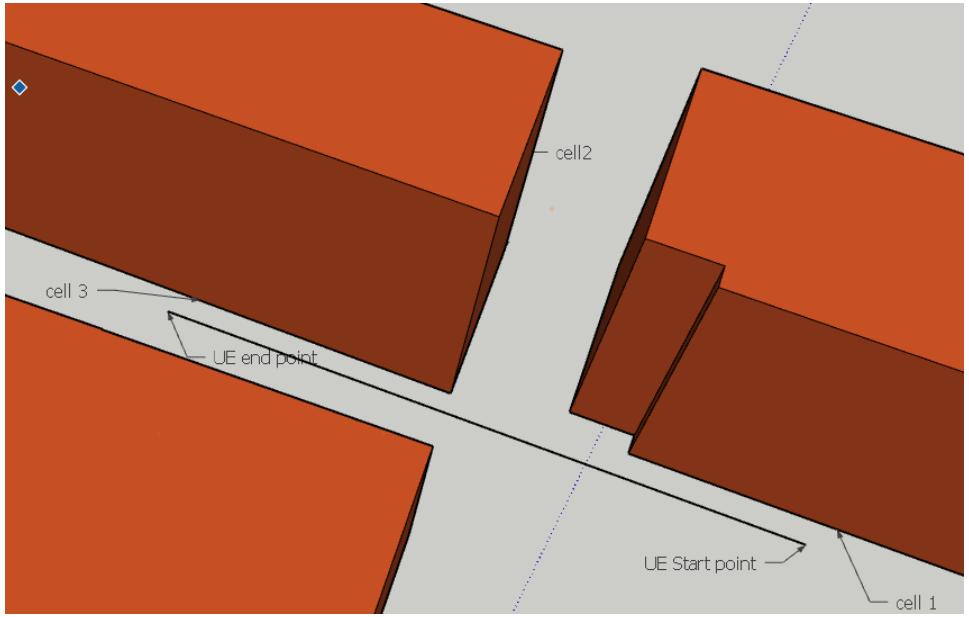


- **Predict RSRP**

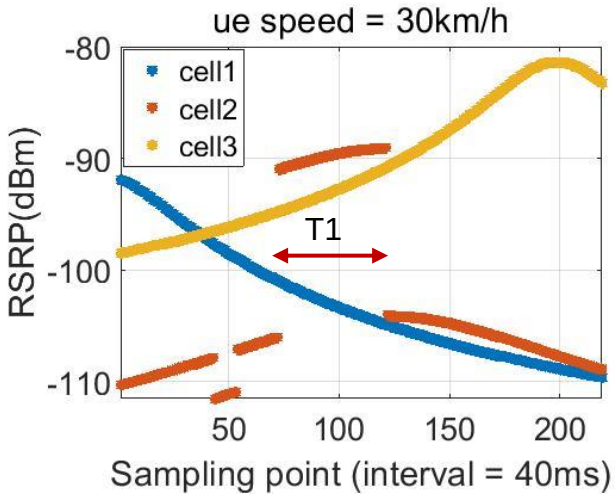


- **UE speed: 80~120km/h**
- **Dataset**
  - Training dataset: 15 UE trajectory
  - Test dataset: 1 UE trajectory
- **AI Model**
  - Fully Connected Neural Network
- **Input**
  - History RSRP of serving & neighbor cell (interval = 10ms)
  - History UE location and speed (interval = 1s)
  - Observation window = 2s
- **Output**
  - Predict RSRP of serving cell
  - Predict RSRP of neighbor cell
  - Prediction timing = 2s
- **Performance**
  - RMSE = 0.8dBm

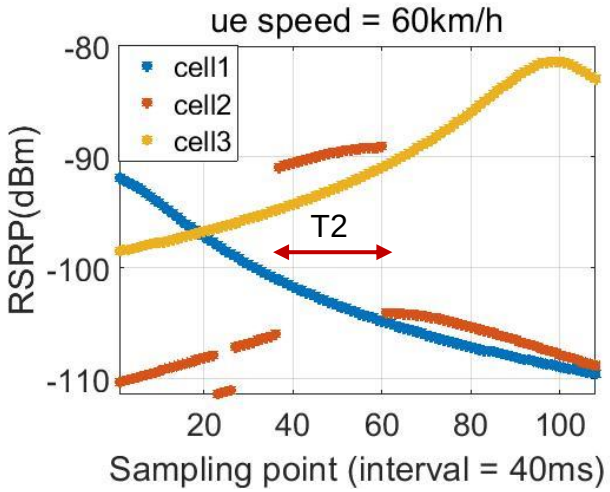
Appendix2: AI+Mobility Ray tracing test



Simulation scenario and UE trajectory



UE1 speed: 30km/h
T1 duration: 1960ms



UE2 speed: 60km/h
T2 duration: 960ms

- When the UE passes through the crossover, the RSRP of cell2 will change dramatically and UE handover to cell 2.
- For UE1 at a low speed, it will be served by cell 2 for a longer time (over 1 second).
- For UE2 at a high speed, it will only be served by cell 2 for less than 1 second.
- Both UEs may experience RLF when leaving the crossover and will reestablishment RRC connection on cell 3.

Appendix2: AI+Mobility Ray tracing test

vivo

■ **Accuracy of RRM measurement prediction**

	Cell 1	Cell 2	Cell 3
Prediction 1	RMSE = 0.0044dB	RMSE = 1.08dB	RMSE = 0.26dB
Prediction 2	RMSE = 0.0062dB	RMSE = 1.11dB	RMSE = 0.26dB
Prediction 3	RMSE = 0.0844dB	RMSE = 1.23dB	RMSE = 0.28dB

Carrier Frequency: FR2, 30GHz

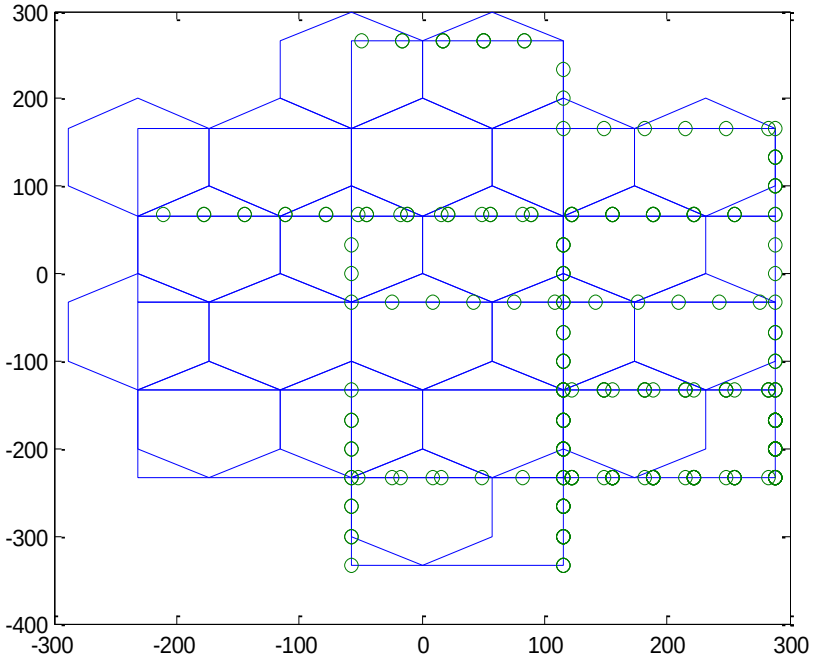
Prediction 1: RSRP of every 80ms in 320ms after T_0

Prediction 2: RSRP of 1s after T_0

Prediction 3: RSRP of 2s after T_0

Appendix2: AI+Mobility System evaluation test

■ Simulation scenario



Simulation scenario and UE trajectory

■ Simulation assumption

Attributes	Values or assumptions
Carrier Frequency	FR1: 4GHz; FR2: 30GHz
TRP Number	7 sites, 3 sector per site
Channel Model	3D-Uma in TR 38.901, support Spatial consistency ISD = 200m
UE speed	120km/h
Mobility management	Event: A3; Hysteresis: 2dB; Offset: 1dB; TimeToTrigger: 320ms, 40ms Handover preparation time: 50ms; Handover execution time: 40ms
RLM	L1 measurement period: 20ms Qin sliding window length: 100ms Qout sliding window length: 200ms Qin threshold: -6dB; Qout threshold: -8dB N310: 1; N311: 1; T310: 1s
Handover model and corresponding metrics	As defined in TR 36.839 Short time of stay: served by the target cell for less than 1s after HO

Appendix2: AI+Mobility System evaluation test

vivo

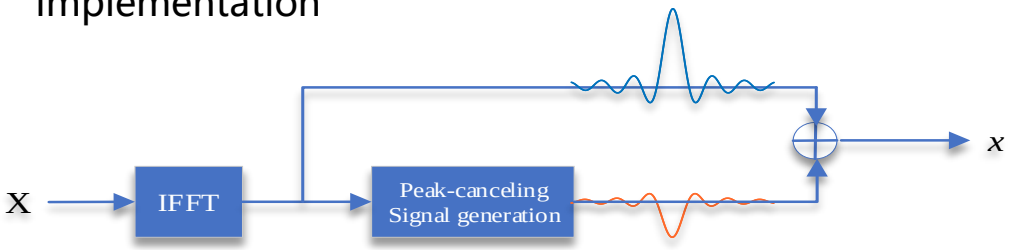
■ **RRM prediction based HO**

		Legacy HO, TTT = 320	Legacy HO, TTT = 40	AI based HO	CHO, TTT = 320	CHO, TTT = 40	AI based CHO
FR1	HOF rate	9.16%	2.2%	1.95%	0.28%	0.15%	0.32%
	Ping-pong HO rate	1.1%	3.6%	0.37%	1.0%	3.7%	0.37%
	Short Time of Stay (1s) rate	13.4%	18.9%	5.7%	13.6%	18.8%	5.67%
FR2	HOF rate	7.4%	2.5%	2.0%	0.42%	0.43%	0.44%
	Ping-pong HO rate	5.2%	10.3%	2.7%	5.2%	10.3%	2.7%
	Short Time of Stay (1s) rate	24.1%	36.7%	10.4%	24.4%	36.5%	10.8%

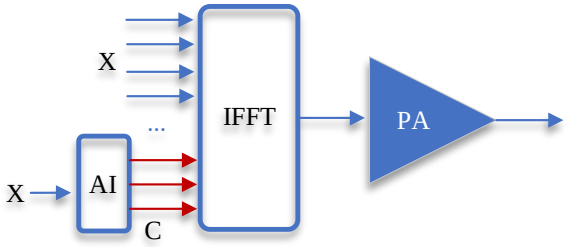
❑ **With RRM prediction, the unintended events rate during HO and CHO can be significantly reduced, including HOF rate, ping-pong HO rate and short time of stay rate.**

Appendix3: AI + Tone Reservation for PAPR reduction

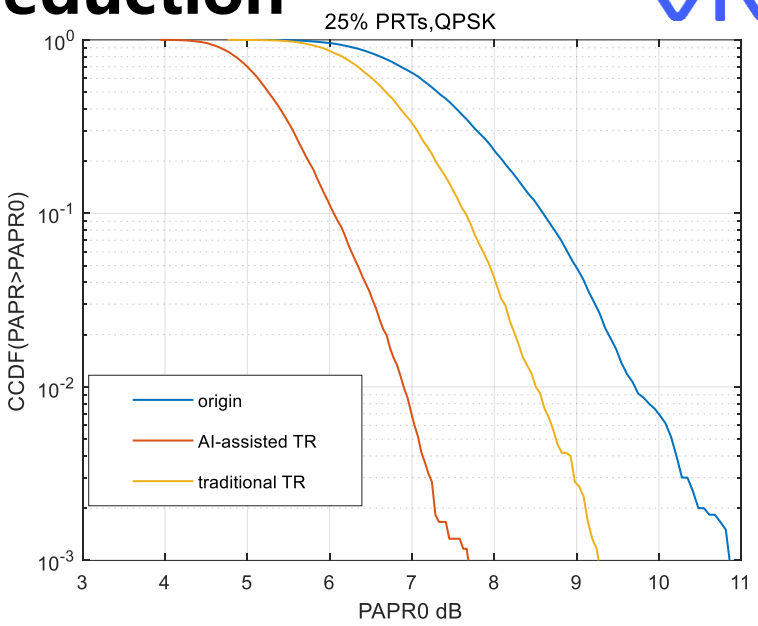
- TR (Tone reserve)
  - Implementation

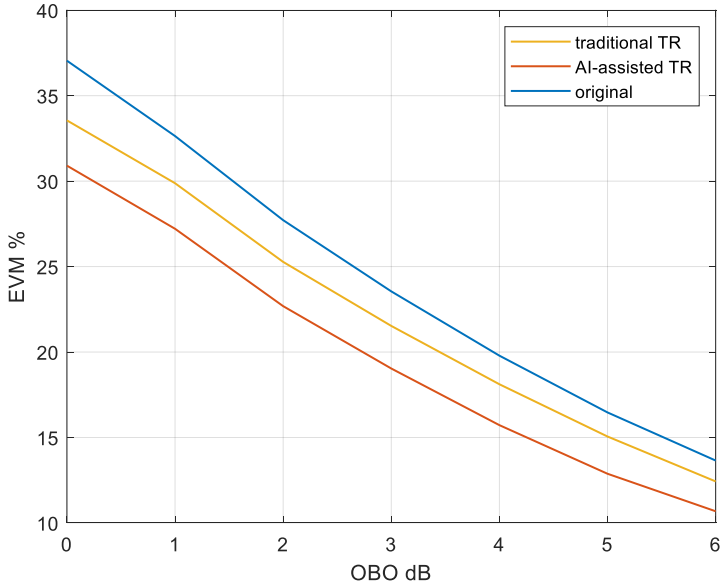


- AI-assisted TR
  - AI-inferenced signal transmitting on reserved tones
  - AI-assisted peak canceling signa generation



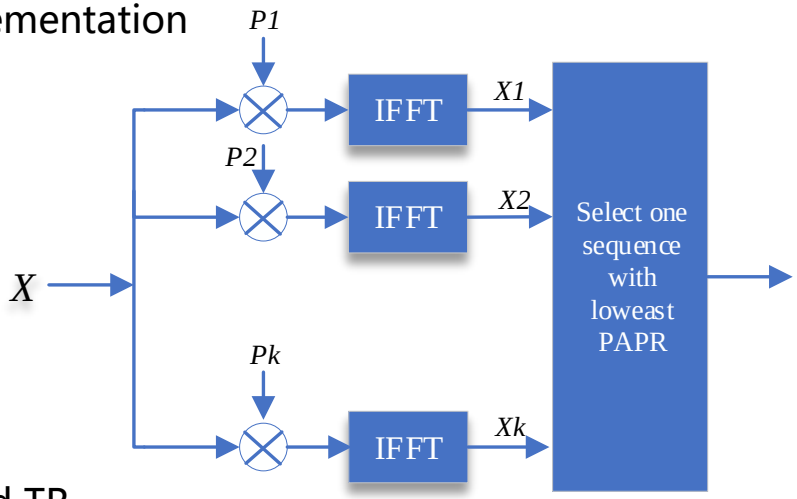
- Potential spec impact
  - Position and number of reserved tones should be informed to receiver

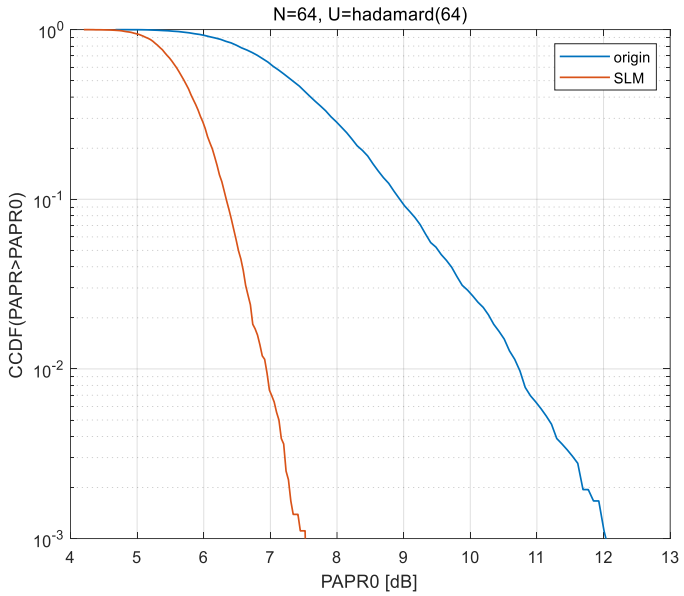




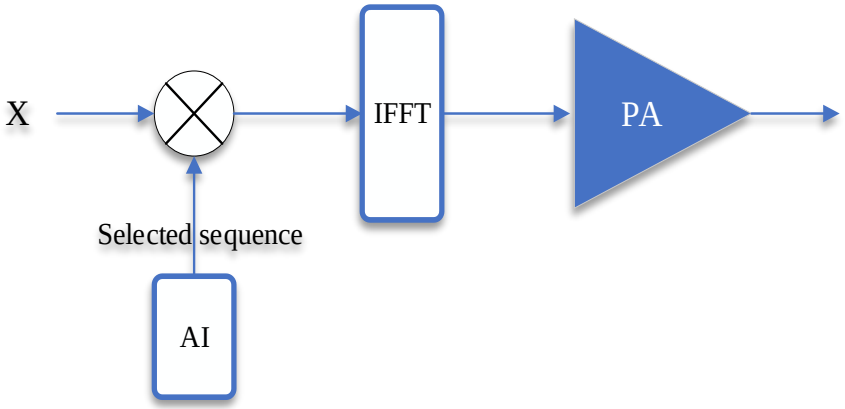
Appendix3: AI + SLM for PAPR reduction

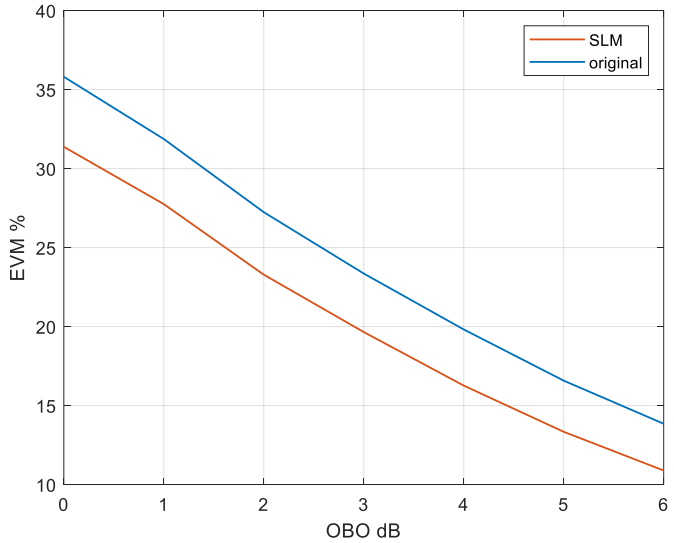
- SLM (selective mapping)
  - Implementation





- AI-assisted TR
  - AI-assisted sequence generation for selective mapping



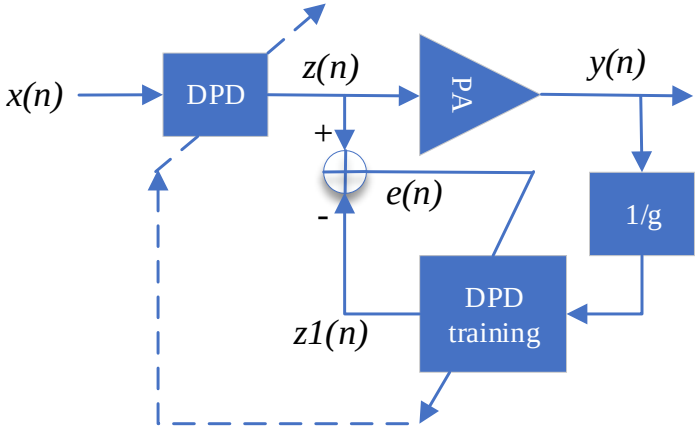


- Potential spec impact
  - Selected scrambling sequence should be informed to receiver for descrambling

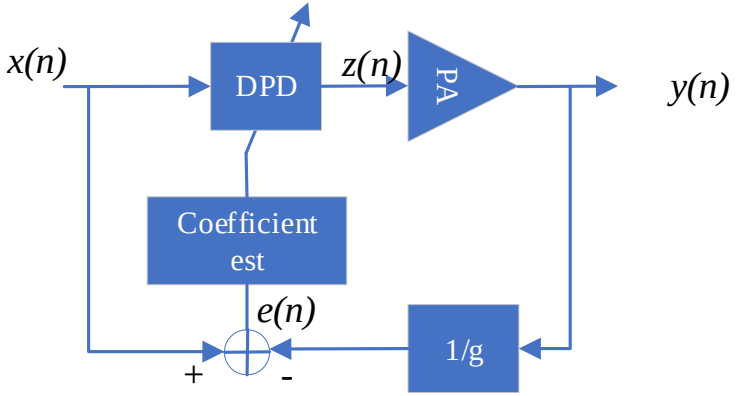
Appendix3: AI + DPD for PA nonlinearity optimization

- DPD
  - Direct learning and indirect learning are considered
  - Performance evaluation

Schemes	NMSE	EVM	ACLR left	ACLR right
w/o DPD	4.7 dB	171%	-35.4 dB	-37.2 dB
Conventional DPD	-35.2 dB	1.73%	-41.2 dB	-42.3 dB
AI-based DPD	-46 dB	0.65%	-46.1 dB	-47.1 dB



Indirect learning



Direct learning

- Potential spec impact is required to achieve adaptive DPD with feedback loop
  - Potential UL transmission delay due to DPD
  - Specific signal transmission for fine tuning with worse EVM/ACLR than required
  - etc.