

3GPP TSG RAN #93

RP-212011

Electronic Meeting, September 13th – September 17th, 2021

Views on study on AIML based air interface enhancement in Rel-18

Source: vivo

Document for: Discussion & Decision

Agenda Item: 9.0.2

General thinking of AI/ML application on air interface

Limitations of current design in air interface

Independent optimization for modules dependent on each other

- E.g., Independent optimization for DL channel acquisition and DL precoding;

Linear sub-optimal algorithms and solutions for non-linear problems

- E.g., Linear channel estimation for non-linear parameter estimation problems

Performance degradation with practical impairments or restrictions

- E.g., EVM degradation and increased unwanted emission when approaching saturated output power due to PA non-linearity;

Non future-proof design without enough flexibility to conduct scenario specific optimization/evolution

- E.g., Performance degradation with type I/II codebook for irregular antenna array in space limited scenarios;
- E.g., Complicated (if not impossible) retuning of system parameters to adjust for evolving traffic/environment



AI/ML is a powerful tool to address challenges for air interface designs

Joint design of several modules could be easily done using AI

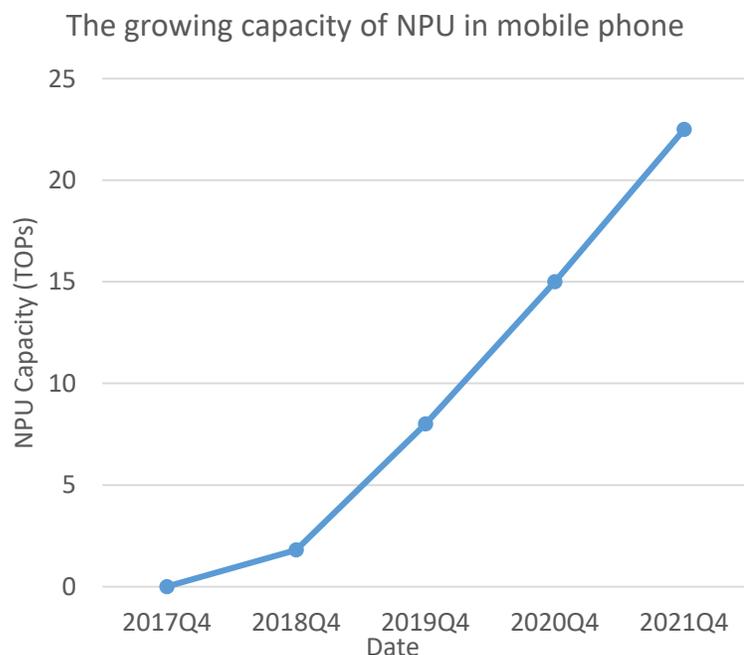
AI solutions could obtain near-optimal solutions for non-linear problems

Modeling of practical impairments becomes possible with AI

PHY layer design based on neural network provides the possibility to continuously evolve for various scenarios

General thinking of AI/ML application on air interface

- AI/ML based air interface is feasible using chipset in current mobile phone :
 - The capacity of one typical NPU used in current mobile phone is 22.5T operations (OPs) per second. From 2017Q1, the capacity of typical NPU in mobile phone is growing very fast year by year.
 - The complexities of typical AI networks are listed in the below table and it is seen that the complexity of AI/ML based air interface is already affordable now.



| | Complexity (OPs) | Ratio of capacity of the typical chipset in 1 second |
|--|------------------|--|
| Inception V2 | 4.1G | 1.8e-4 |
| Inception V3 | 12G | 5.3e-4 |
| CaffeNet | 724M | 3.2e-5 |
| GoogLeNet | 2G | 8.9e-5 |
| MobileNet | 1.15G | 5.1e-5 |
| AI network for DMRS in the slides | 2.9M | 1.3e-7 (1.3e-4 in 1ms) |
| AI network for CSI feedback in the slides | 8.9M | 4.0e-7 (4.0e-4 in 1ms) |
| AI network for beam management in the slides | 544K | 2.4e-8 (2.4e-5 in 1ms) |
| AI for channel prediction in the slides | 7.3M | 3.2e-7 (3.2e-4 in 1ms) |

General thinking of AI/ML application on air interface

- AI/ML based technology is data driven, with its livelihood lying in evolution based on data from practical engineering problems;
 - 3GPP provides the best platform for such data driven evolution;
- Rel-18 is an important release that may put the basis for future two or more releases;
- Study of potential areas for AI/ML application on air interface should be started in Rel-18;

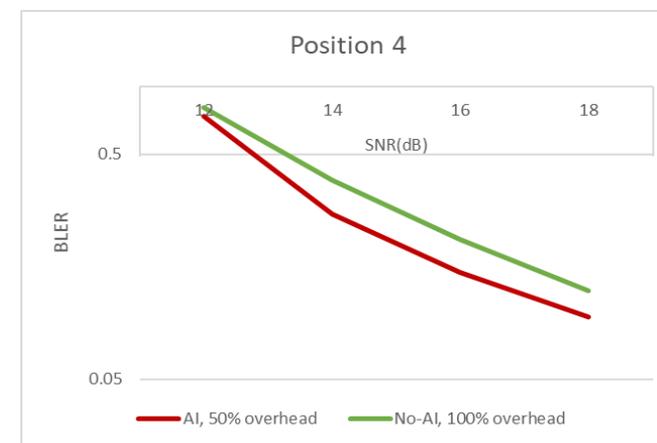
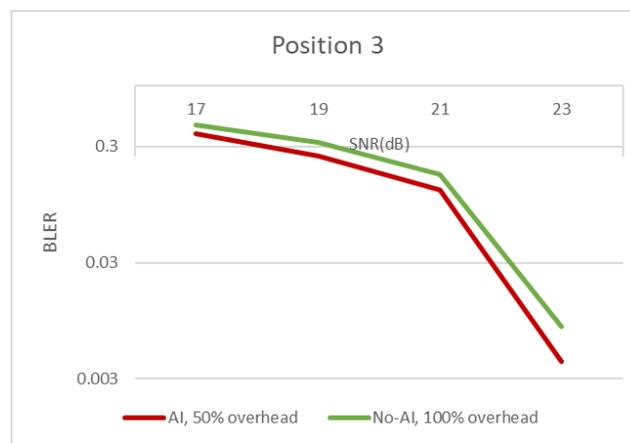
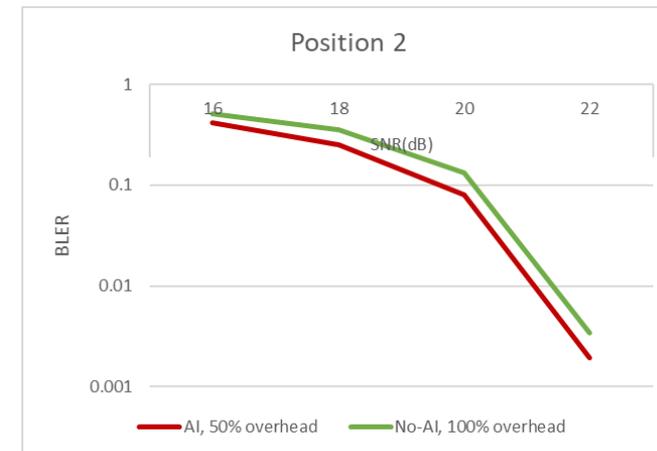
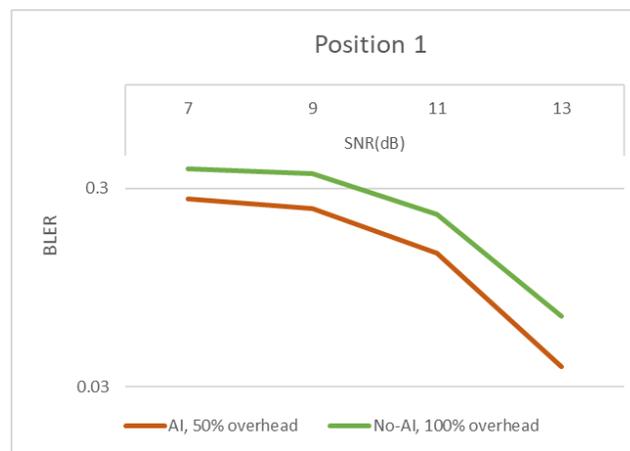
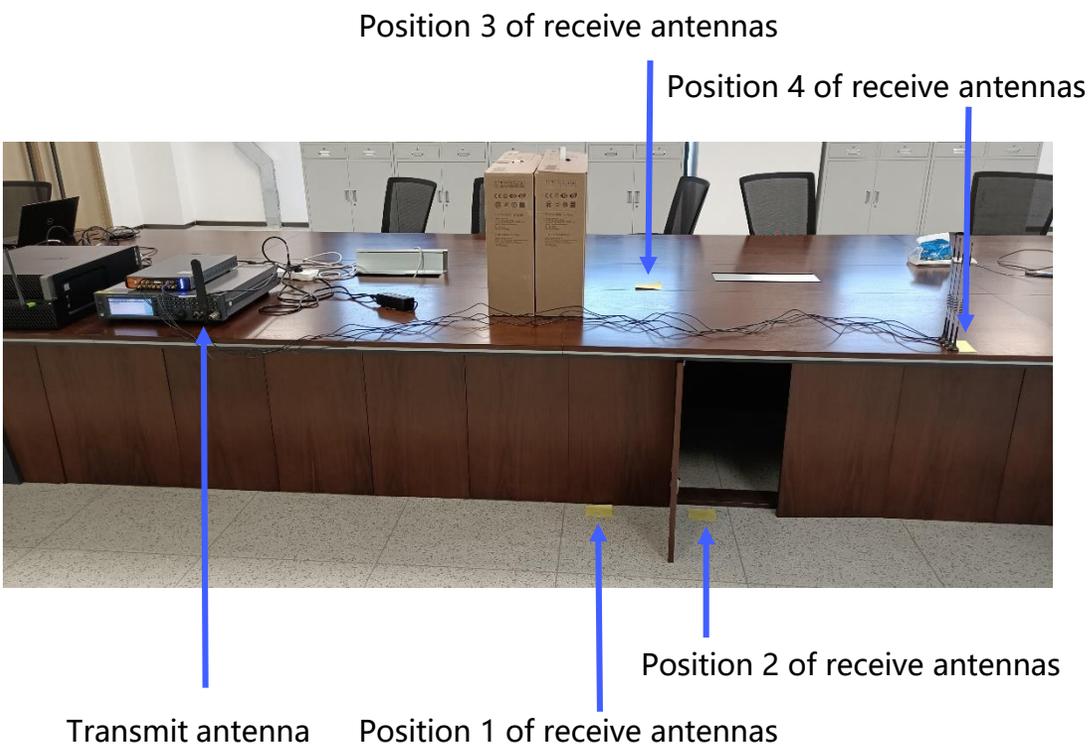


Potential Areas for application of AI/ML on air interface

- Almost every building block of communication link can apply AI/ML to improve performance. The following examples from initial study shows performance gains (Appendix and next page)
 - CSI feedback with lower overhead to achieve higher system efficiency
 - Positioning performance improvement under various challenging conditions
 - RS overhead reduction and channel estimation performance improvement, including DMRS/CSI-RS/SRS
 - Intra-cell/Inter-cell beam management latency and accuracy improvement
 - Channel prediction to improve system performance in high mobility scenarios
 - Mobility enhancement to improve UE experience during handover
- AI/ML can be applied at either one side or both sides of the communication link:
 - Application of AI/ML at UE side, e.g., channel estimation with new QCL chains, with AI/ML receivers at UE side;
 - Application of AI/ML at network sides, e.g., lower RS overhead, including UL DMRS, SRS, with AI/ML channel estimation at gNB side;
 - Areas with AI/ML application at both sides of network and UE, e.g.,
 - CSI report enhancement, with AI/ML to encode UCI at UE side and decode UCI at gNB side;
 - Positioning enhancement, with AI/ML at UE side to extract features that would be used at the network side for positioning related estimate, or vice versa;
- Both training and inference of AI/ML should be considered for the applied areas:
 - Learning of radio environment through e.g., RS transmission, CSI report;
 - Aggregation of learning from multiple nodes for the applied area;

AI/ML+DMRS channel estimation: over the air experiments

- The near-optimal AI/ML network largely reduces the MSE, without assistance of TRS, with half overhead compared to Rel-15 NR design.
- The model is trained with limited LLS generated data set, but still show performance gains when directly applied over the air.

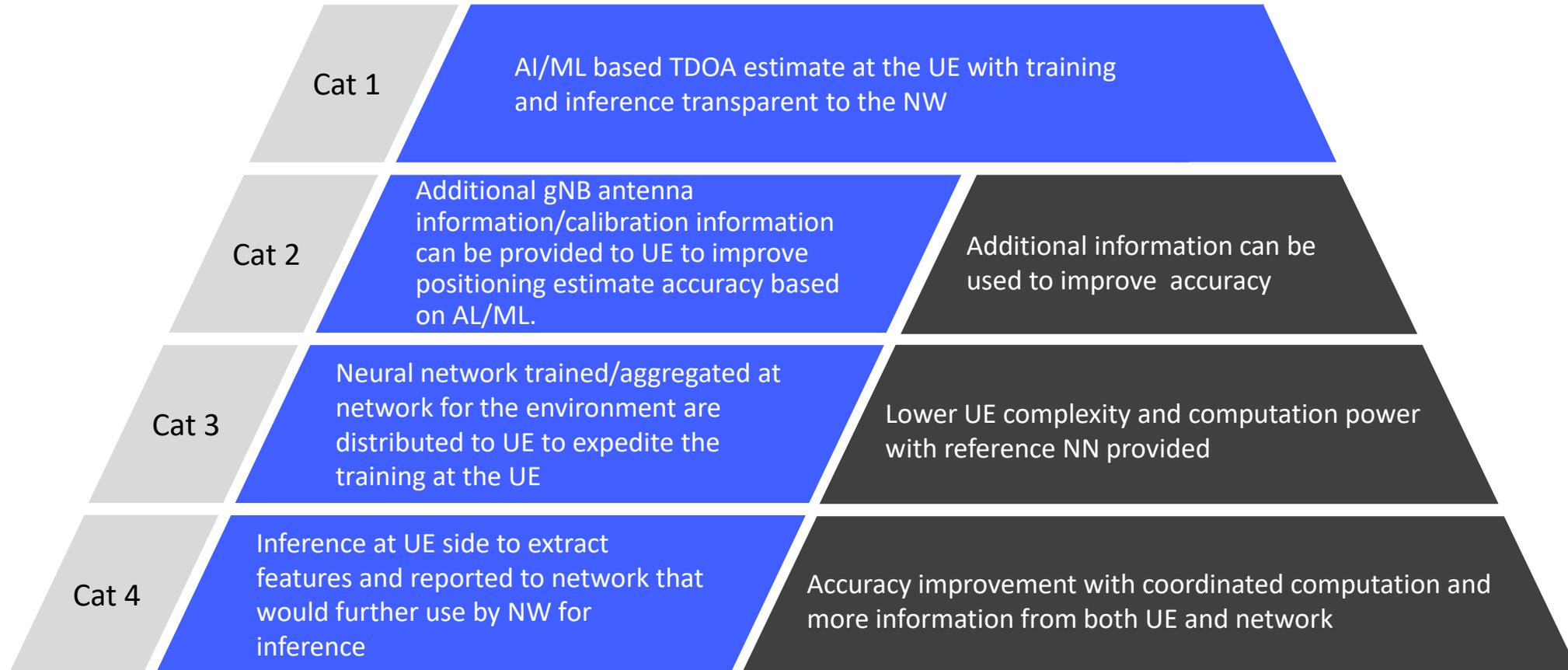


Parameters: One AI model is used for all positions. This AI model is trained by using LLS channel data, in which SNR 0~34dB, UE speed 0~1km/h, TDL model with random delay and random power for each path, maximum delay spread 0~200ns, carrier frequency 3.6GHz, 1 gNB antenna, 4 UE antennas, 4 RBs for AI model input and 48 RBs for air test. DMRS takes 6 subcarriers in 1 RB and 2 symbol in 1 slot.

Categorization of use cases

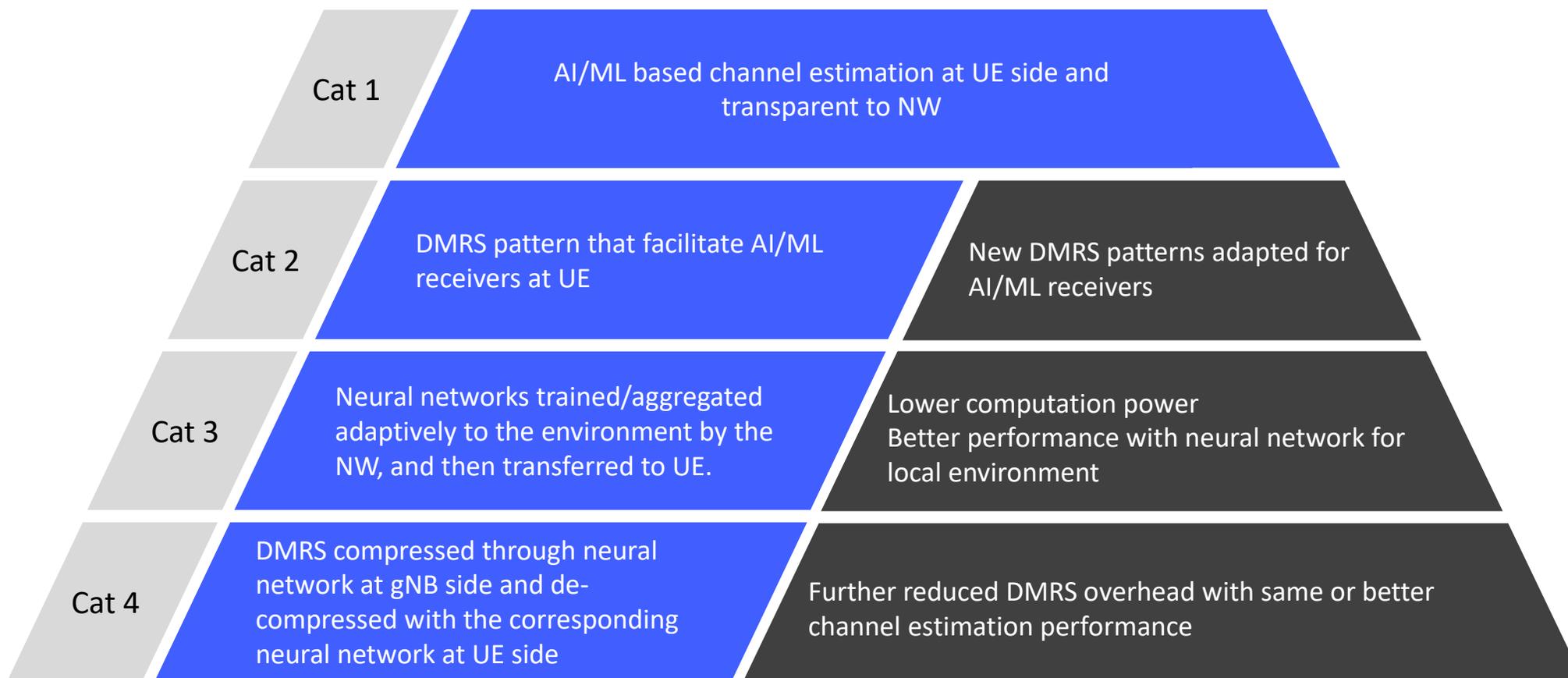
- Use cases can be categorized from system impact perspective:
 - Cat1: AI/ML related training and inference are all conducted at one side of network or UE and is transparent to the other side;
 - Cat2: AI/ML related training and inference are conducted at one side of network or UE, but requires additional signaling or procedure enhancements between two sides, potentially with existing signaling framework.
 - Additional information is not directly related to training and inference, e.g., capability, new patterns etc.;
 - Training or inference is done by implementation and not explicitly seen in the specification
 - Cat3: AI/ML related inference is conducted at one side of network or UE, with assistance information exchanged between two sides;
 - E.g., neural network models provided as a reference for UE or for aggregation;
 - Cat4: AI/ML related inference are conducted together at both sides of network and UE
 - training maybe conducted at one side or both
 - Information related to inference need to be exchanged between both sides;

Example#1: Different categories for AL/ML based positioning



Positioning

Example#2: Different categories for AI/ML based channel estimation



Channel Estimation

Potential use cases that would benefit from AL/ML

- For each use case, different categories may apply
 - In general, Cat N would have better performance than cat N-1, but also more system impact, including spec efforts
 - Expected system performance and implied specification work would be different for different categories.
 - A pre-study phase is needed to identify and categorize the potential sub-use cases for each use case, including details of each sub-use case , e.g., typical models used, training details, UE and network involvement etc.

| | SRS | TRS | CSI-RS | DMRS | Positioning | CSI | Beam | Mobility |
|-------|-----|-----|--------|------|-------------|-----|------|----------|
| Cat 1 | √ | √ | √ | √ | √ | √ | √ | √ |
| Cat 2 | √ | √ | √ | √ | √ | √ | √ | √ |
| Cat 3 | | √ | √ | √ | √ | √ | √ | √ |
| Cat 4 | | | √ | √ | √ | √ | √ | √ |

- Data set for training and test can be constructed based on statistical model in 38.901
 - Data set constructed with 38.901 could effectively emulate local radio environment and also representative of various practical channel conditions;
 - Large number of samples can be generated for training, testing and verification with 38.901;
- Generalization performance should also be considered when constructing verification data set, e.g.,
 - Large scale parameter perturbation when generating verification data set;
 - System level model in 38.901 should mainly be considered for verification;
- Model alignment between companies
 - Fixed model for calibration between companies
 - Selected and recommended models for evaluation for applied areas based on company input
 - Reported models from companies for evaluation results capturing

Potential SI/WI timeline and scope

- Pre-study phase (6 months): Identify and categorize use case(s) for AI/ML based air interface enhancements
 - Collection of AI/ML background information for each use case, including information for data set construction, typical neural network models, model training details, UE and network involvement in the use case and etc.
 - Typical deployment scenarios and KPIs for each use case
 - Categorization of use cases based on collaboration frameworks between UE and network
 - Down-selection of use cases for performance evaluation phase
- Performance evaluation phase (12 months) :
 - Performance evaluation of the use cases concluded in pre-study phase, including
 - Study of evaluation methodology, including data set construction, model alignment between companies, etc
 - Evaluation assumptions and performance comparison for each use case
 - Framework and procedures of applying AI/ML for the use cases concluded in Pre-study phase, including potential specification impact;
- Work item phase to be started based on the output from the WG study item in later releases

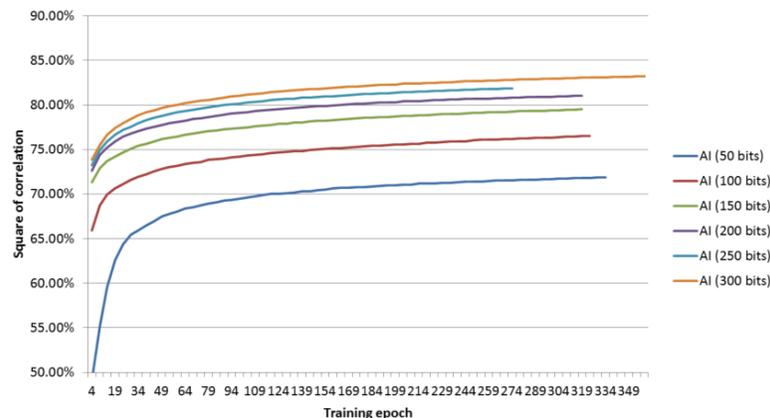
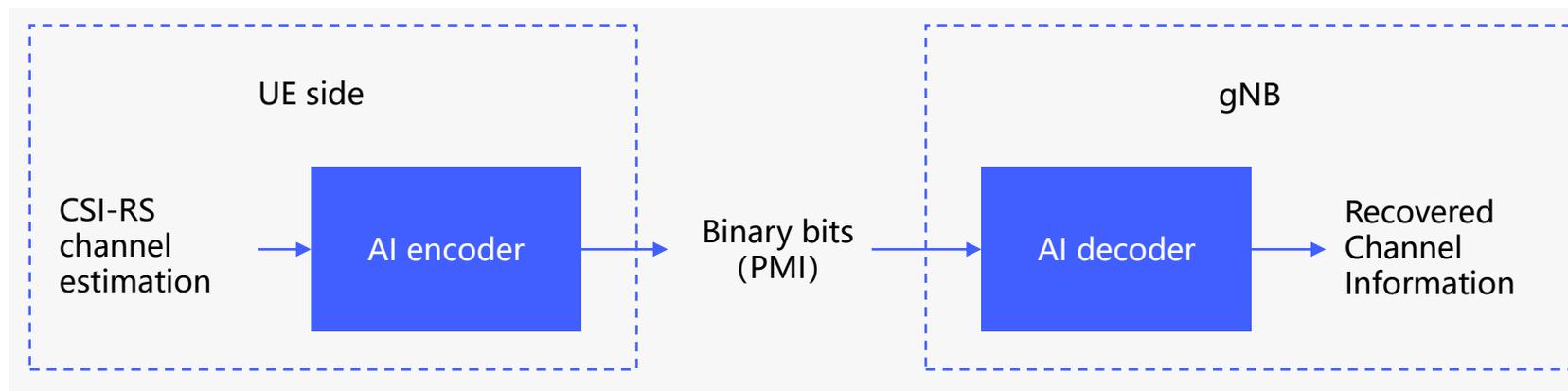
THANK YOU.

谢谢。

Appendix with Example Cases

Example Case1: AI/ML+CSI feedback

- For low overhead case, 30% Tput gain is achieved using AI, compared to NR specified solutions.

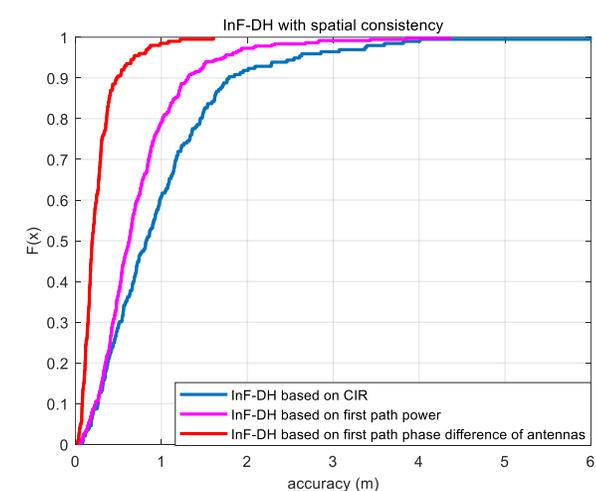
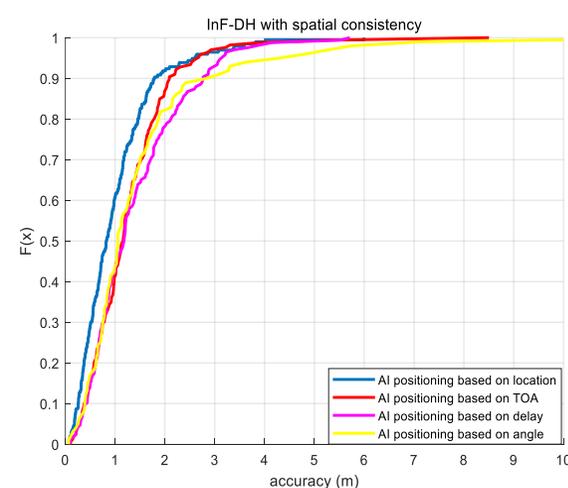
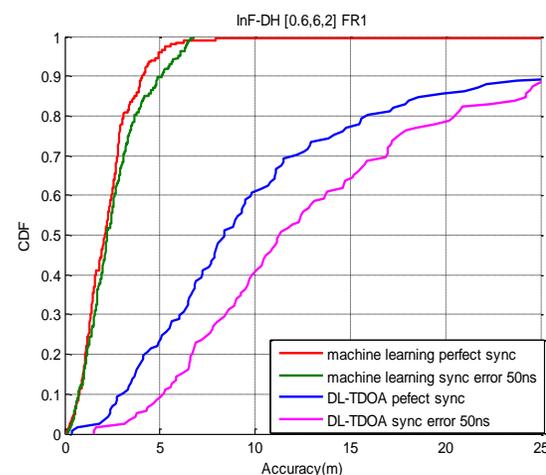
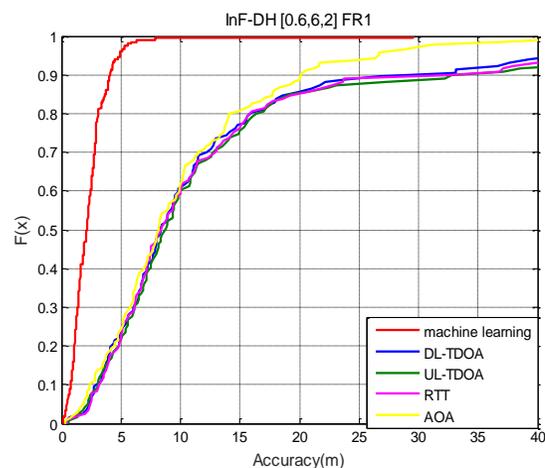


| Method | Spectrum efficiency (bit/s/Hz) |
|-----------------------|--------------------------------|
| NR specified solution | 6.41 |
| AI | 8.28 |

Parameters: SLS, UMi 38.901, 7 cells, 3 sectors for each cell, UE speed 3km/h, carrier frequency 3.5 GHz, 32 gNB antenna ([Mg Ng M N P] = [1 1 2 8 2]), 4 UE antenna ([Mg Ng M N P] = [1 1 2 2 2]), 52 RBs. The overhead of PMI is 58 bits.

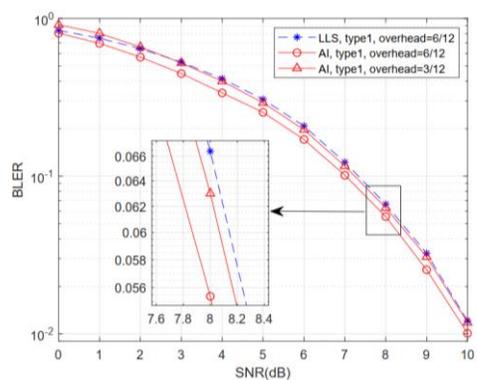
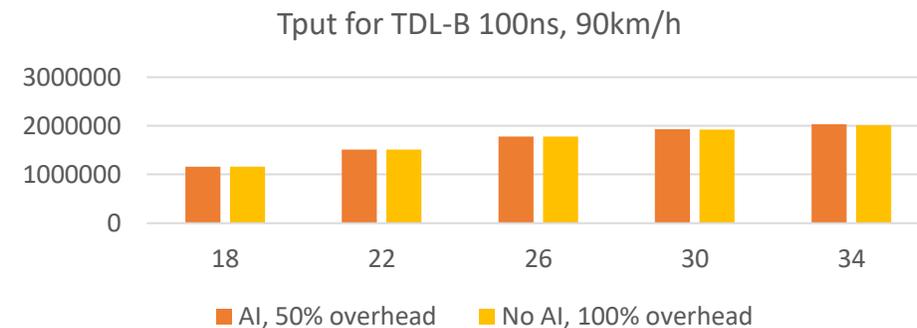
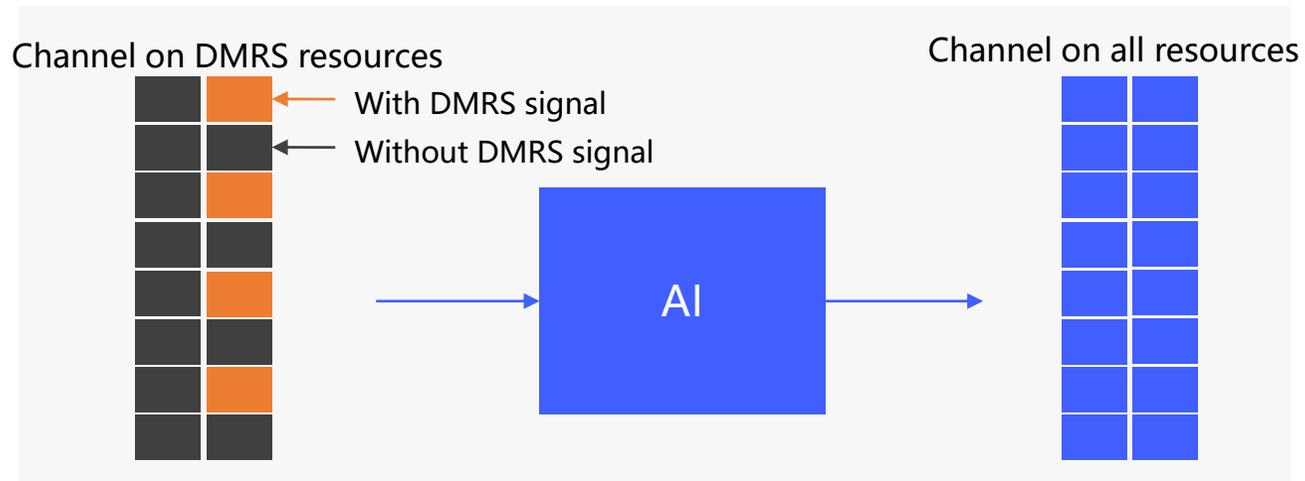
Example Case2: AI/ML + Positioning

- Compared to Rel-16/Rel-17 solutions, AI/ML based solutions could increase the accuracy dramatically at least from the following aspects
 - To combat NLOS
 - To combat synchronization error between different TRPs
 - To combat Tx-Rx timing error

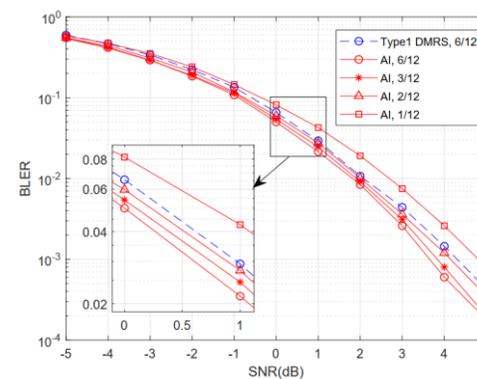
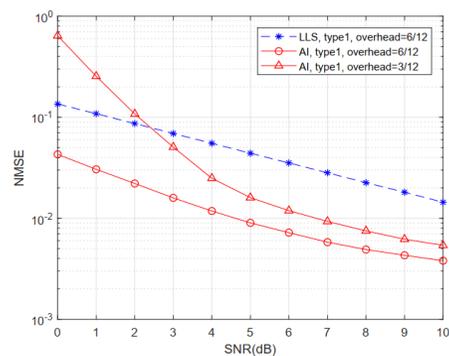


Example Case3: AI/ML+DMRS channel estimation

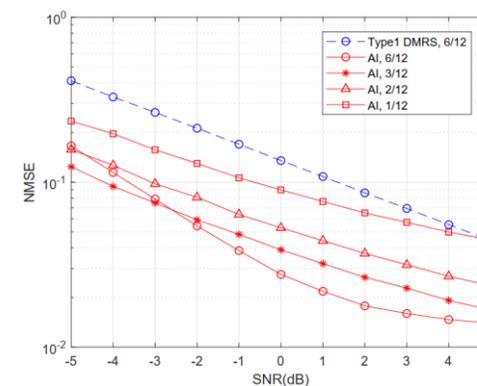
- The near-optimal AI/ML network largely reduces the MSE with lower DMRS overhead and without assistance of TRS.



TDL-A, type1, MCS7



TDL-A, type1, MCS0



Example Case3: AI/ML+DMRS channel estimation, experiments with channel emulators

- The near-optimal AI/ML network largely reduces the MSE, without assistance of TRS, with half overhead compared to Rel-15 NR design.



Channel emulator



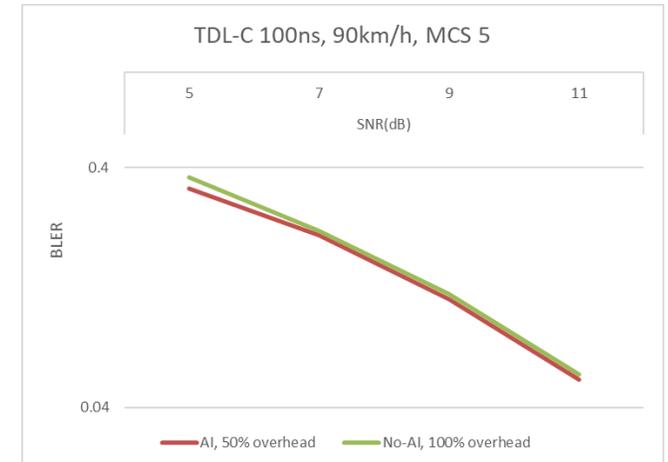
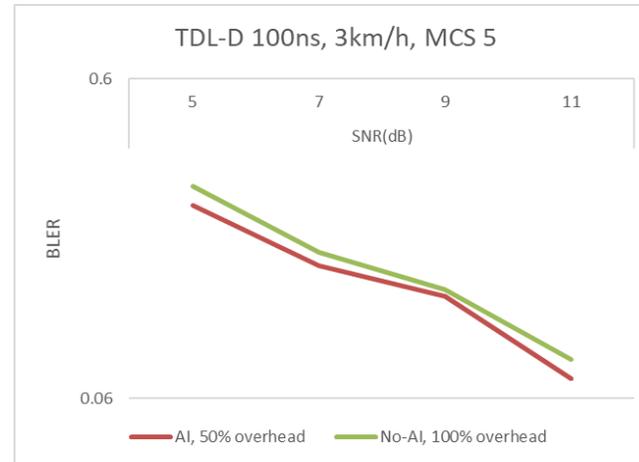
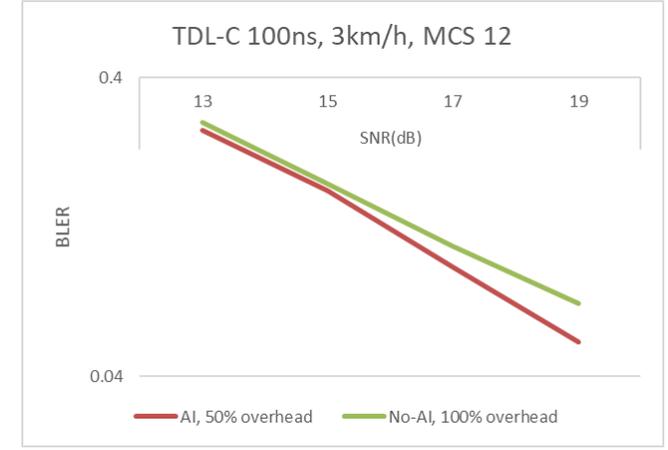
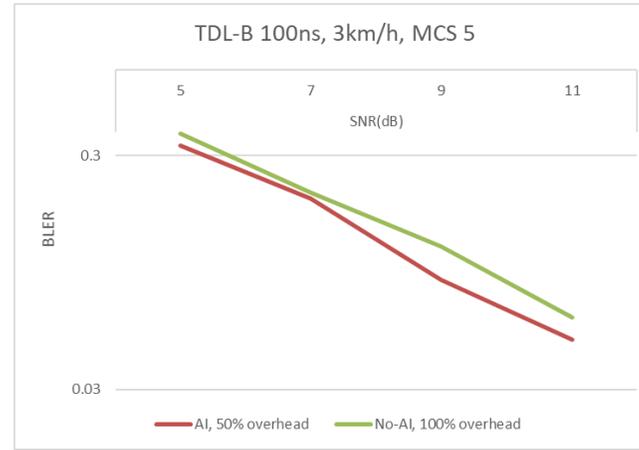
X86+FPGA



Signal transeiver (USRP)



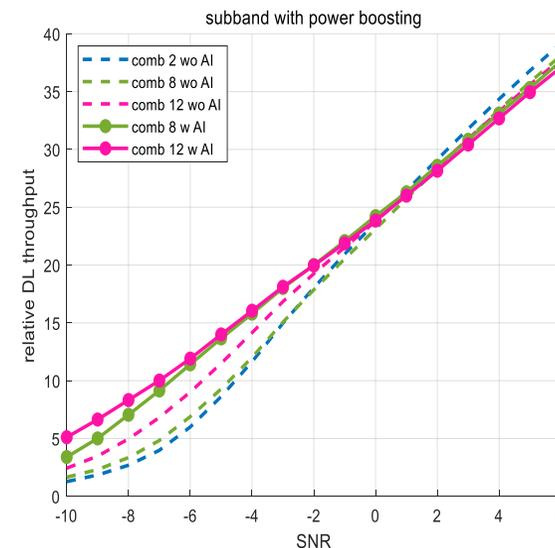
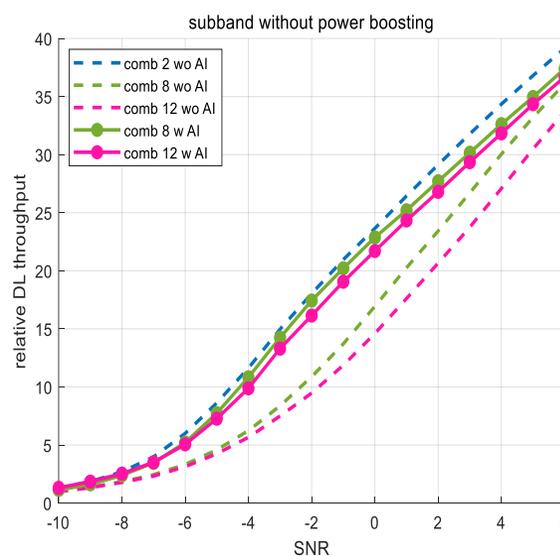
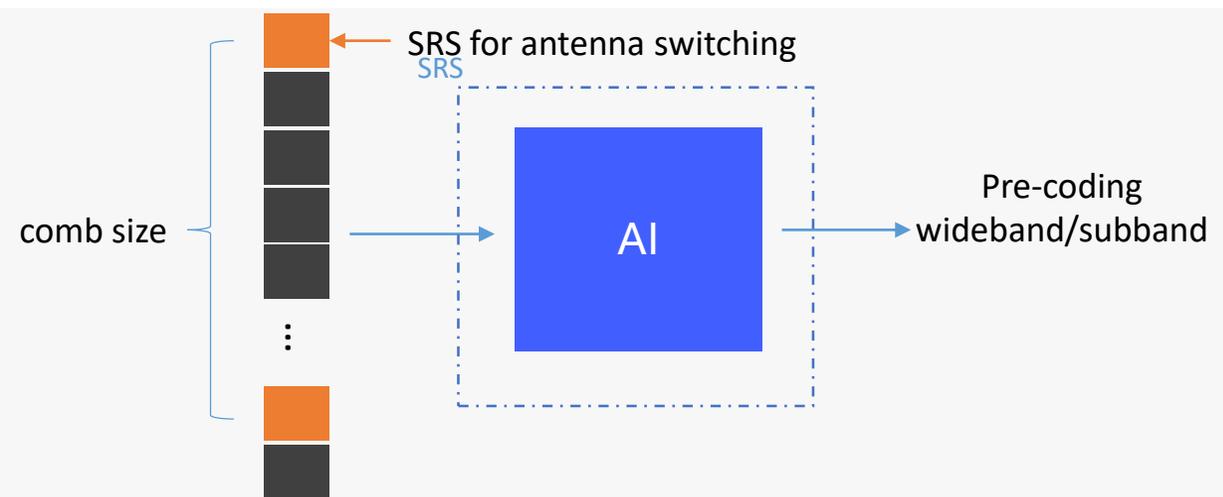
Signal generator



Parameters: One AI model is used for all scenarios. SNR -5~30dB, UE speed 0~90km/h, TDL-A/B/C/D/E, delay spread 0~300ns, carrier frequency 3.6GHz, 1 gNB antenna, 4 UE antennas, 16 RBs. DMRS takes 6 subcarriers in 1 RB and 2 symbol in 1 slot.

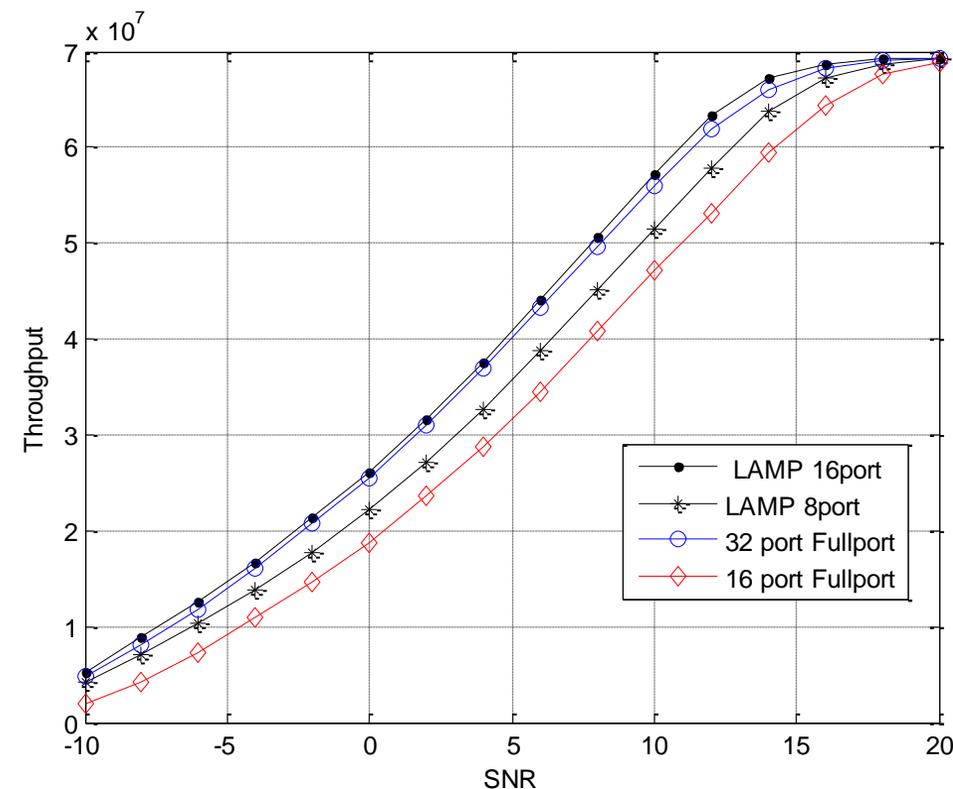
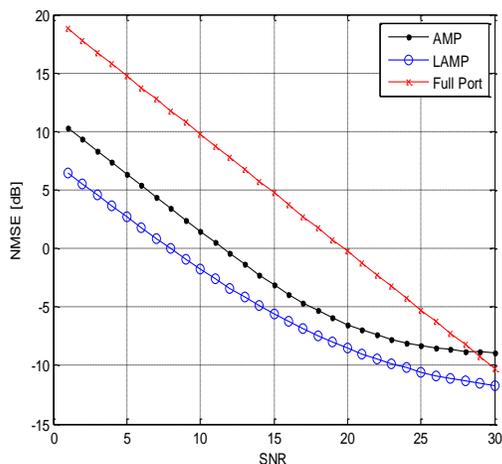
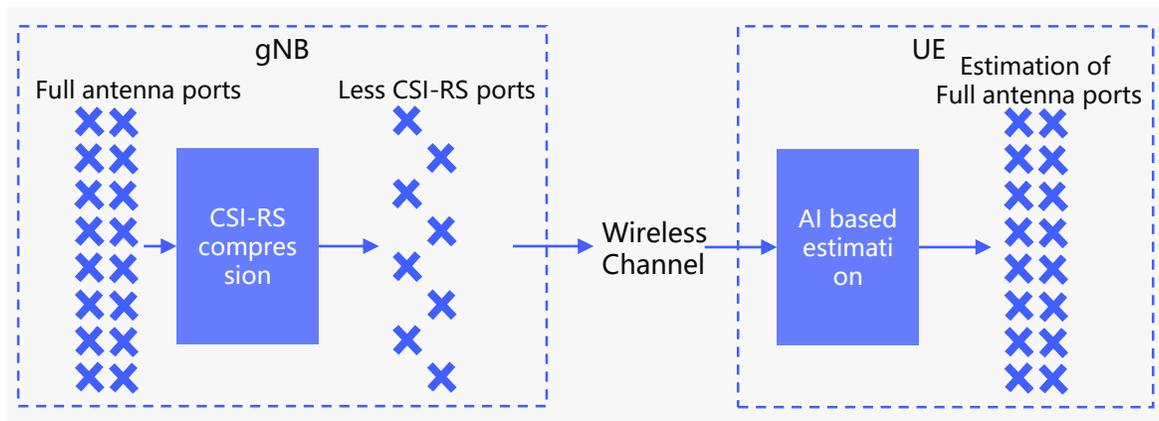
Example Case4: AI/ML+SRS overhead reduction for DL CSI

- In comparison with comb 2 SRS in Rel 15/16,
 - AI-based DL CSI acquisition for comb size = 12 achieves similar throughput performance without power boosting
 - AI-based DL CSI acquisition for comb size = 12 obtains more than 2dB gain in low SNR range if power boosting applied



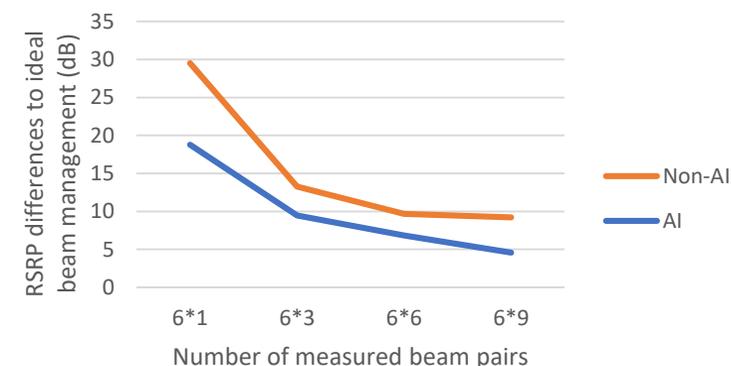
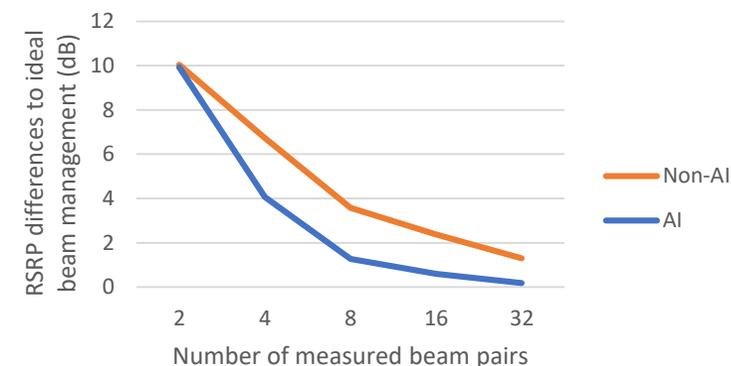
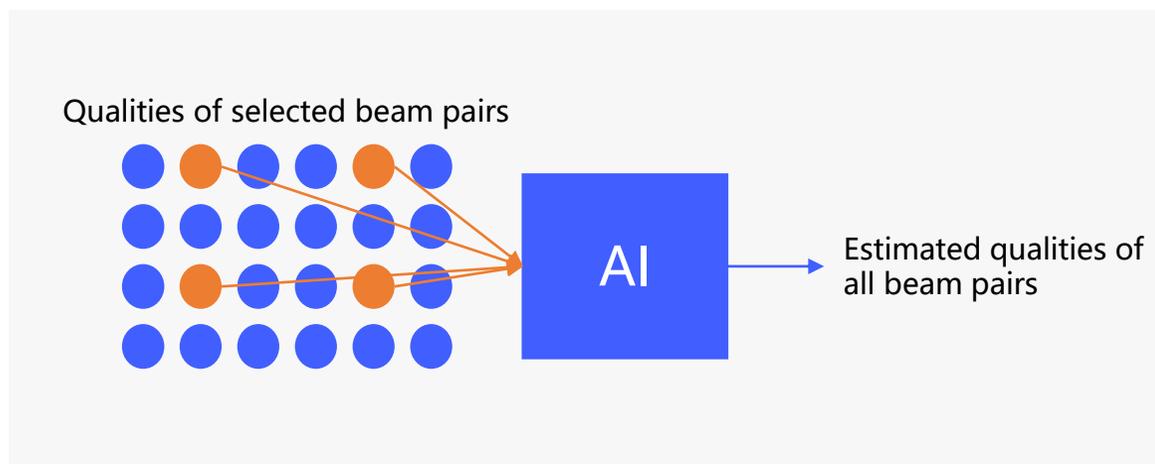
Example Case5: AI/ML+ CSI-RS overhead reduction

- Using 50% CSI-RS overhead, AI based CSI-RS compression could achieve better throughput than traditional algorithm with 100% CSI-RS overhead.



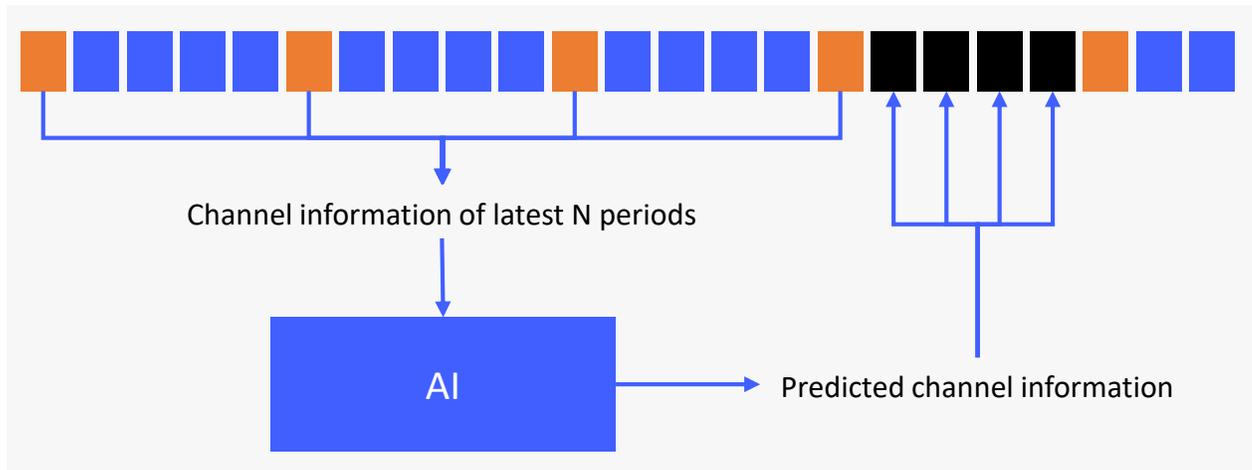
Example Case6: AI/ML + Fast intra-cell/inter-cell beam selection

- Measure small number of beam pairs, and use AI to estimate qualities of all beam pairs. It could used for both intra-cell beam management and inter-cell beam management.
 - Compared to the measurement of 32 beam pairs in non-AI method, AI only needs to measure 8 beam pairs hence reducing measurement time by 75%, expediting overall beam management process.
 - Compared to the measurement of 54 beam pairs in non-AI method, AI only needs to measure 18 beam pairs hence reducing measurement time by 67%, expediting overall beam management process.



Example Case7: AI/ML+ Channel prediction

- Using the outdated channel information from previous RS detection, AI could predict the future channel information very well.



| Case | NMSE | |
|-----------------------------------|-------------|--------------|
| | Speed 3km/h | Speed 30km/h |
| No channel prediction | 0.15 | 1.6 |
| AI with information of 2 periods | 2.5e-8 | 2.6e-7 |
| AI with information of 4 periods | 3.3e-9 | 4.2e-8 |
| AI with information of 6 periods | 7.9e-10 | 1.9e-8 |
| AI with information of 8 periods | 5.0e-10 | 5.6e-9 |
| AI with information of 10 periods | 3.3e-10 | 4.2e-9 |

Example Case8: Mobility enhancement with AI/ML based trajectory prediction

- For RRM measurements and handover procedure, beam tracking/handover decision can be enhanced based on the trajectory prediction of UE.
 - UE can measure less beams to reduce the power consumption
 - gNB can schedule less RS resource to reduce the overhead for beam measurement of the target cell
 - AI functionalities , e.g. model inference, can locate in UE side

