

3GPP TSG RAN Rel-18 workshop

RWS-210170

Electronic Meeting, June 28 – July 2, 2021

Study on AI/ML based air interface enhancement in Rel-18

Source: vivo

Document for: Discussion & Decision

Agenda Item: 4.3

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

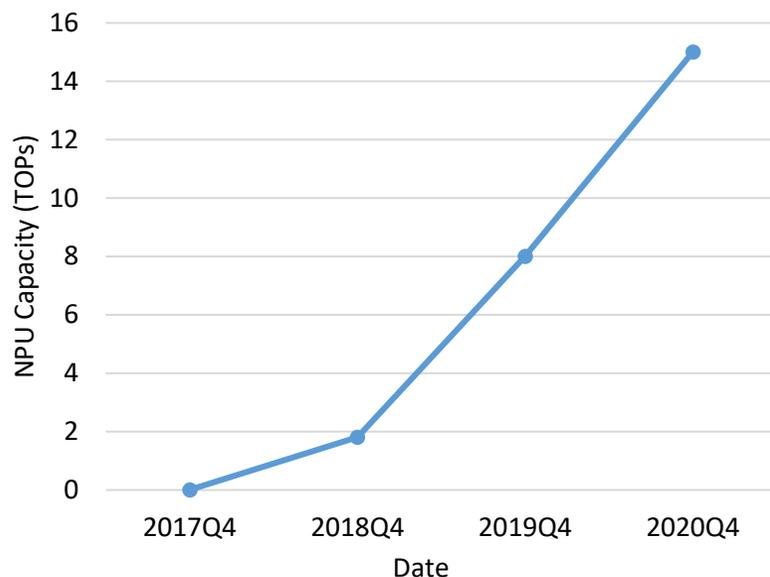
Modeling of practical impairments becomes possible with AI

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 15T 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.

The growing capacity of NPU in mobile phone



	Complexity (OPs)	Ratio of capacity of the typical chipset in 1 second
Inception V2	4.1G	2.7e-4
Inception V3	12G	8.0e-4
CaffeNet	724M	4.8e-5
GoogLeNet	2G	1.3e-4
MobileNet	1.15G	7.7e-5
AI network for DMRS in the slides	2.9M	1.9e-7 (1.9e-4 in 1ms)
AI network for CSI feedback in the slides	8.9M	5.9e-7 (5.9e-4 in 1ms)
AI network for beam management in the slides	544K	3.7e-8 (3.7e-5 in 1ms)
AI for channel prediction in the slides	7.3M	4.9e-7 (4.9e-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;

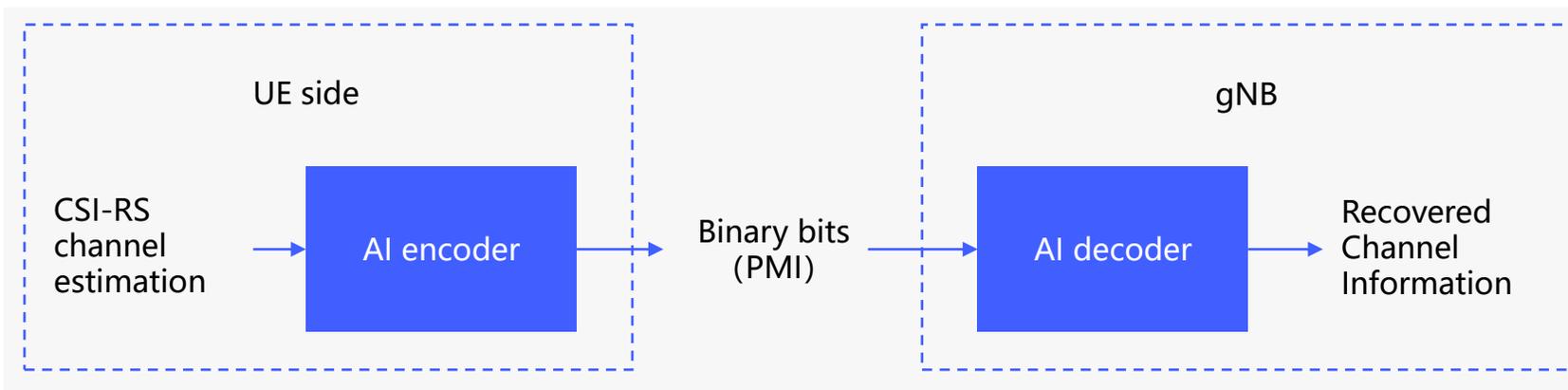


Potential Areas for application of AI/ML on air interface

- Almost every building block of communication link can apply AI/ML to improve performance. The following examples from initial study shows performance gains (Appendix).
 - Positioning performance improvement under various challenging conditions;
 - CSI feedback with lower overhead to achieve higher system efficiency;
 - RS overhead reduction and channel estimation performance improvement, including DMRS/CSI-RS/SRS;
 - Intra-cell/Inter-cell beam management latency and accuracy improvement
- 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;

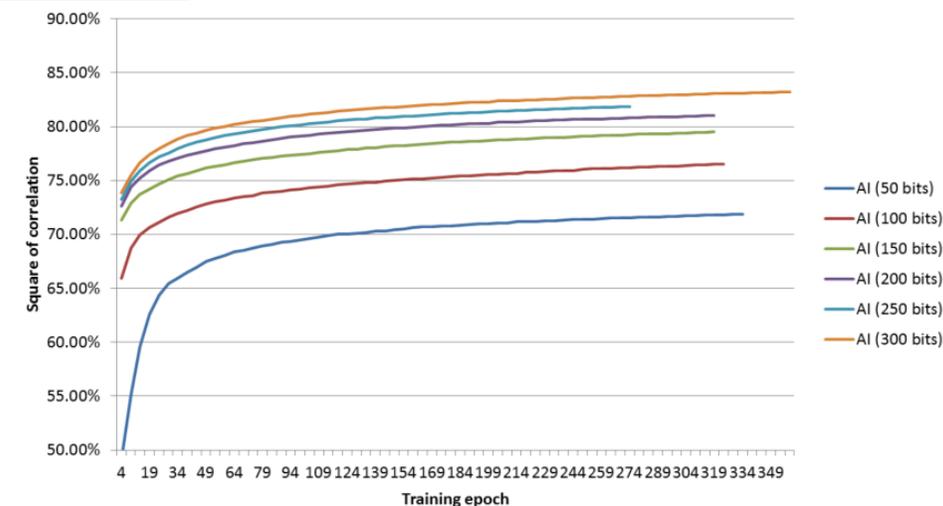
Example Case1: AI/ML+CSI feedback

- AI network is not limited to certain restrictions of current codebook design.
- 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

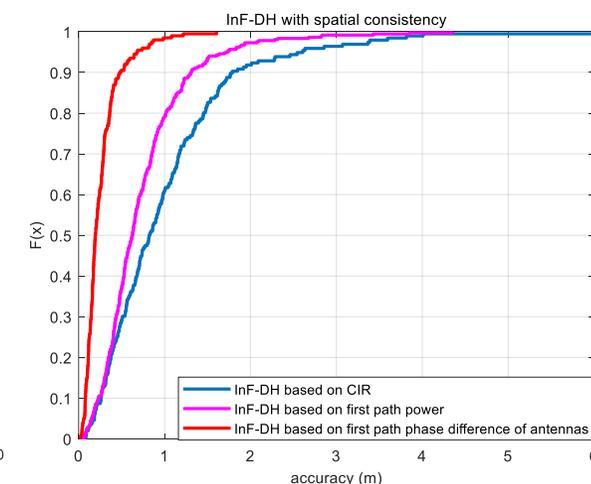
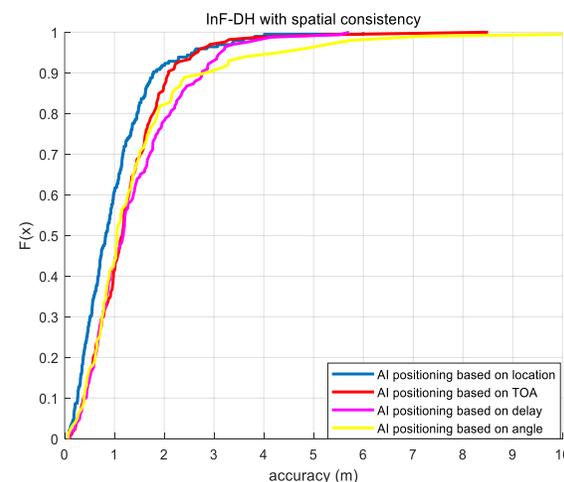
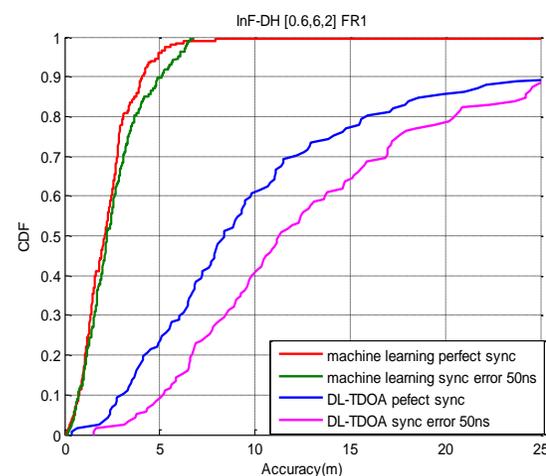
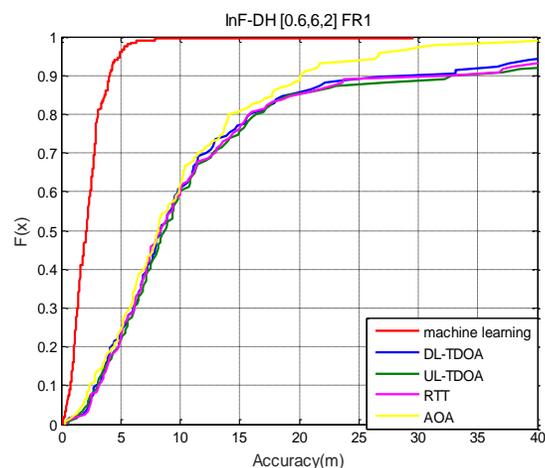
- Potential specification includes at least:
 - Framework and procedures on exchange of neural network related information between network and UE that facilitate learning on radio environment;
 - Exchange of neural network for aligned CSI computation between UE and gNB;
 - New CSI measurement and computation procedures;



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

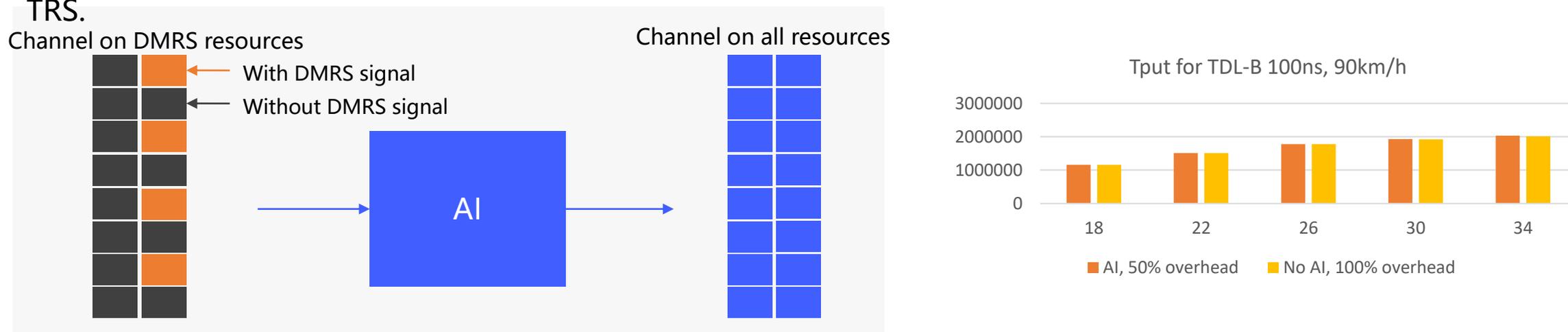
- Compared to Rel-16/Rel-17 solutions, AI/ML based solutions could increase the accuracy dramatically at least from the following aspects
 - To combat NLOS in DH scenarios
 - To combat synchronization error between different TRPs
 - To combat Tx-Rx timing error



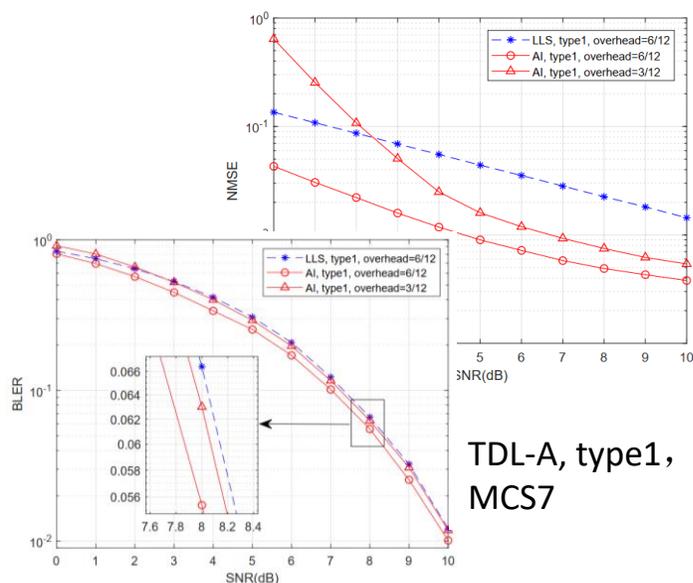
- Potential specification impact includes at least:
 - Framework and procedures on exchange of neural network related information between network and UE that facilitate learning and inference;
 - New measurement and reporting procedures with aligned neural network between gNB and UEs;

Example Case3: AI/ML+DMRS channel estimation

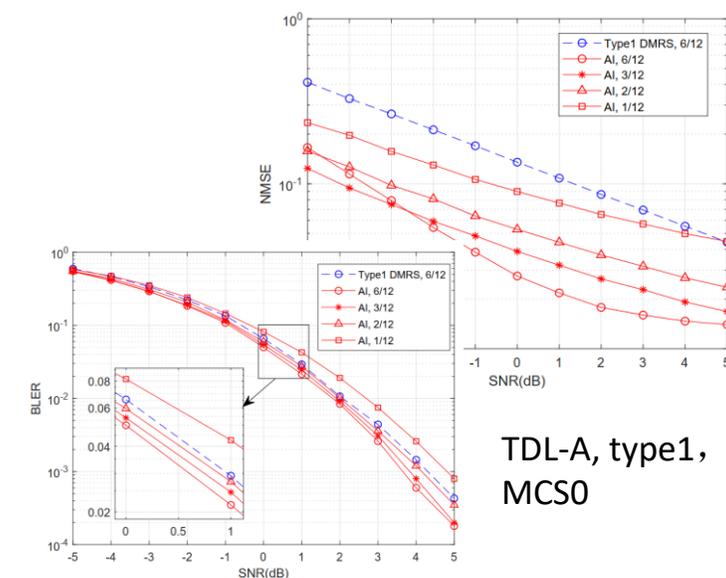
- The near-optimal AI/ML network largely reduces the MSE with lower DMRS overhead and without assistance of TRS.



- Potential specification impact includes at least
 - Exchange of neural network related information between network and UE that facilitate learning on radio environment;
 - QCL relationship and TRS transmission and reception that facilitate the new receivers;
 - DMRS patterns optimized for the new receivers;



TDL-A, type1, MCS7



TDL-A, type1, MCS0

Example Case3: AI/ML+DMRS channel estimation

- 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



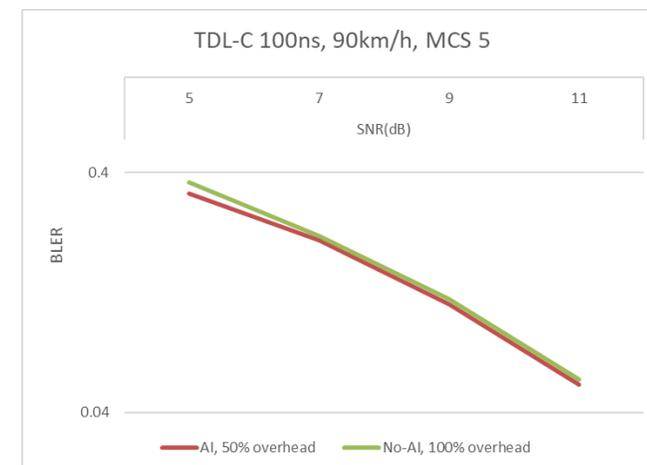
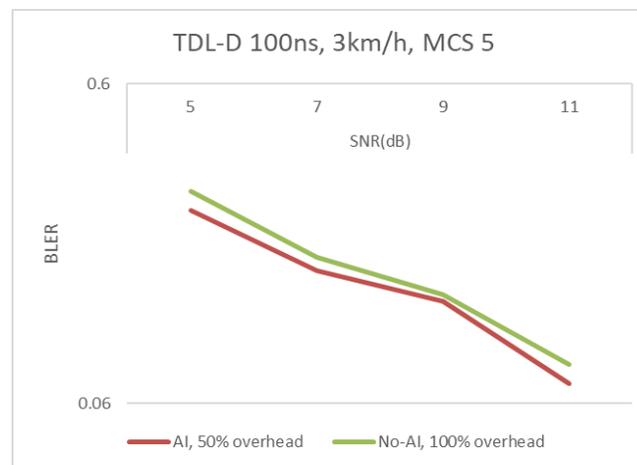
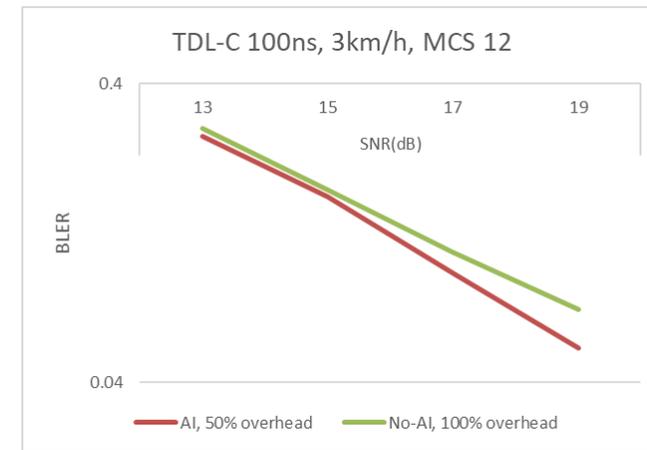
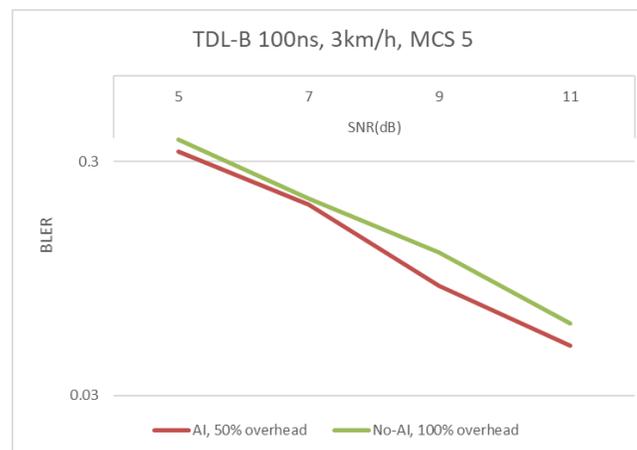
X86+FPGA



Signal transeiver (USRP)



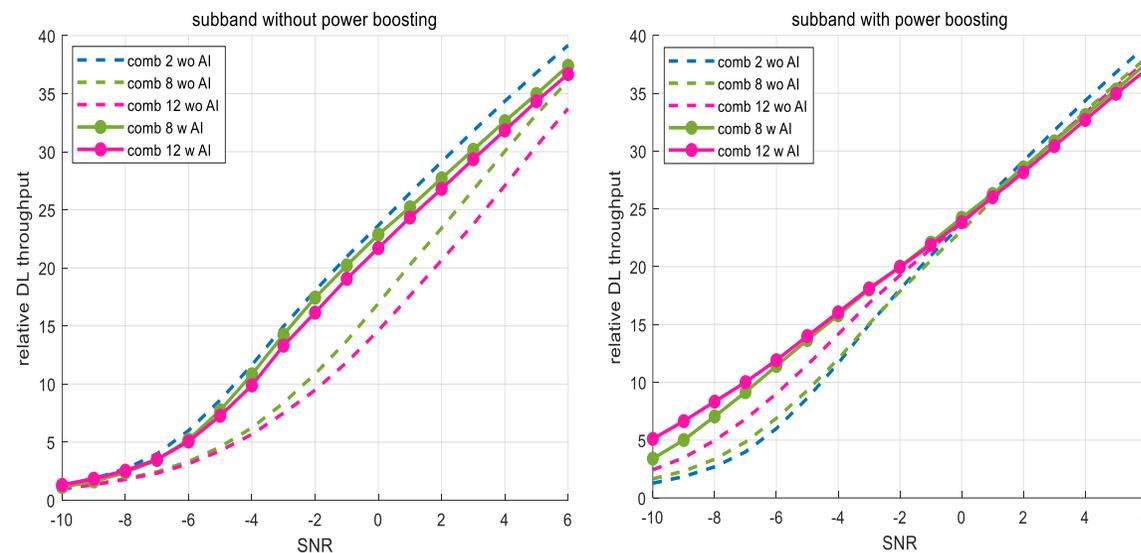
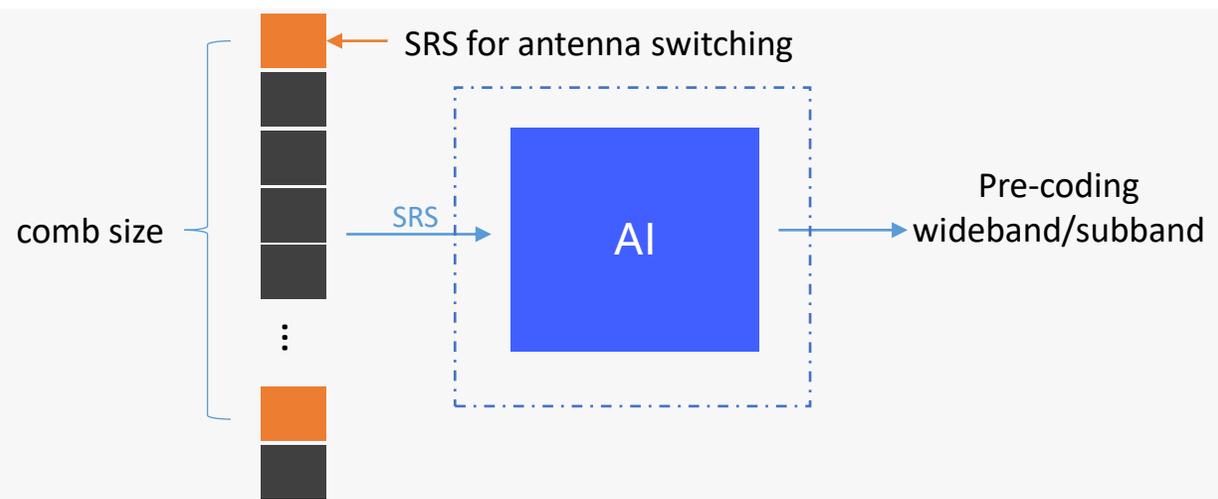
Signal generator



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

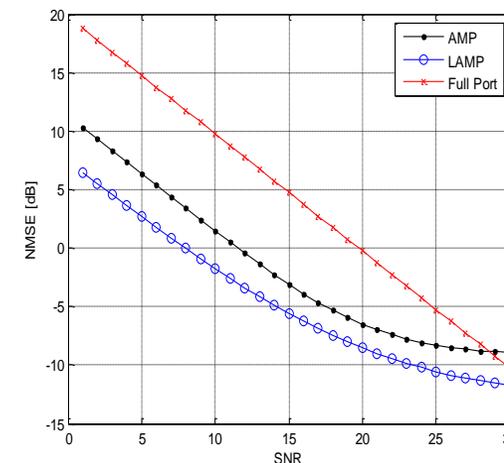
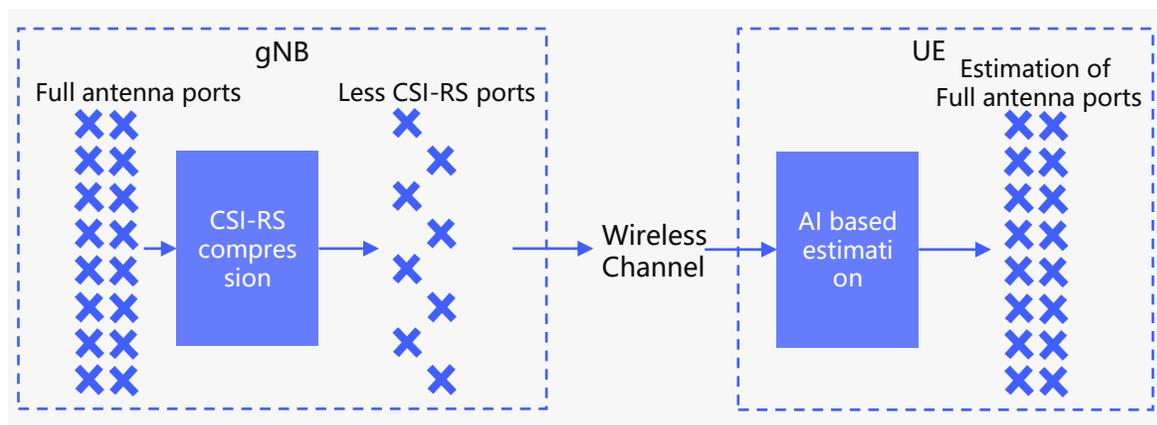
- In comparison with comb 2 SRS in Rel 15/16,
 - AI-based DL CSI acquisition for comb size = 12 achieves similar throughput performance without power boosting
 - AI-based DL CSI acquisition for comb size = 12 obtains more than 2dB gain in low SNR range if power boosting applied



- Potential specification impact includes at least
 - Exchange of neural network related information between network and UE that facilitate learning on radio environment;
 - SRS patterns optimized for the new receivers;

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

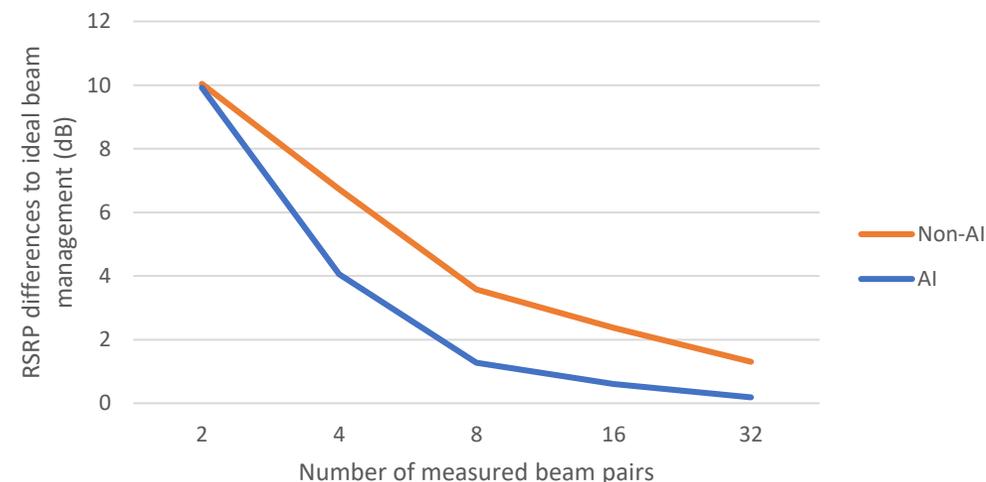
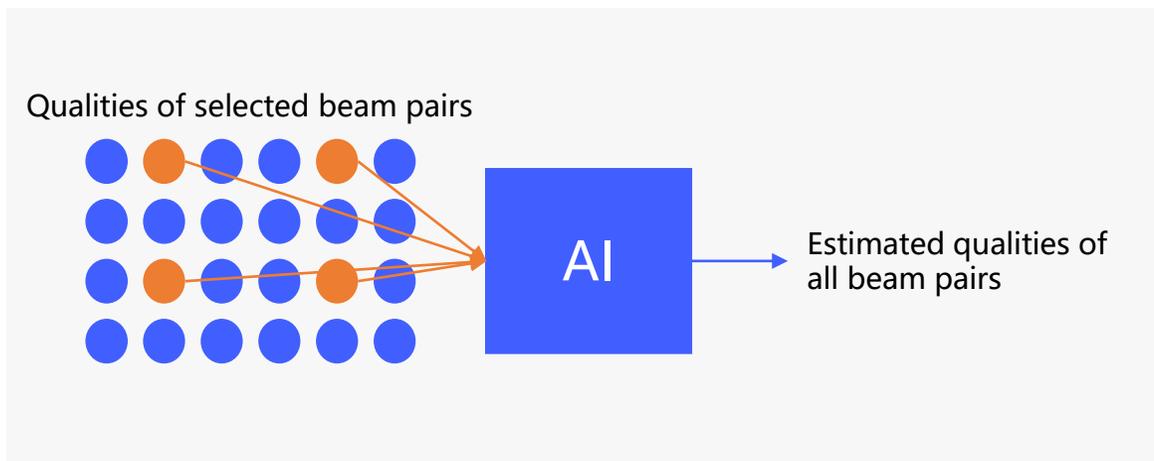
- Using 50% CSI-RS overhead, AI based CSI-RS compression could achieve better throughput than traditional algorithm with 100% CSI-RS overhead.



- Potential specification impact includes at least:
 - Exchange of neural network related information between network and UE that facilitate learning on radio environment;
 - Exchange of neural network that is used for the UE to generate full port CSI;
 - CSI-RS transmission and reception with compressed ports;

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

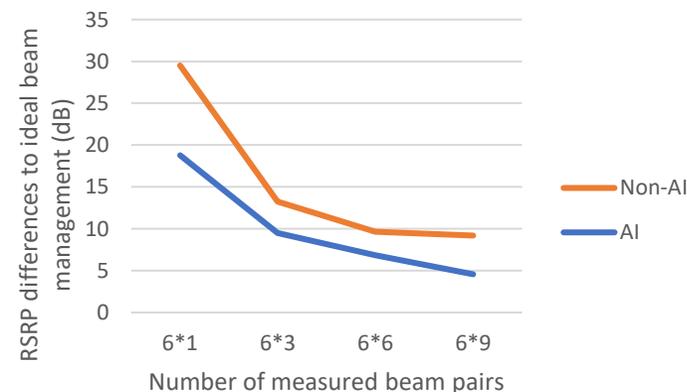
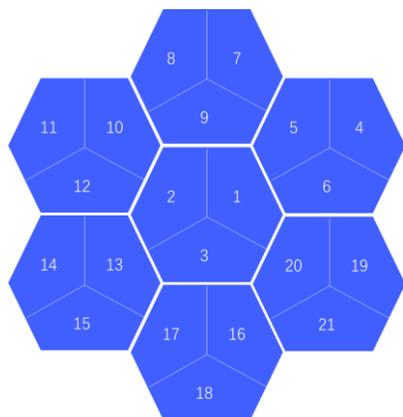
- Measure small number of beam pairs, and use AI to estimate qualities of all beam pairs. It could be used for both intra-cell beam management and inter-cell beam management.
- Compared to the measurement of 32 beam pairs in non-AI method, AI only needs to measure 8 beam pairs hence reducing measurement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



- Potential specification impact includes:
 - Framework and procedures on exchange of neural network related information between network and UE that facilitate learning and inference;
 - Exchange of neural network for aligned beam measurement between UE and gNB;
 - Support of beam measurement reports with more than 4 beams;

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

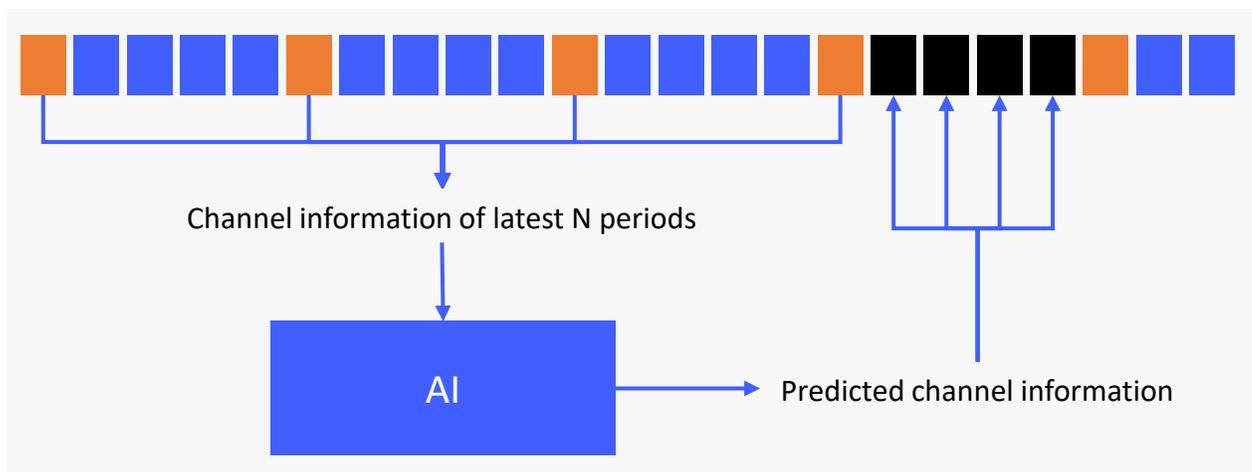
- Similar mechanism could also be used to facilitate inter-cell measurement;
 - Compared to the measurement of 54 beam pairs in non-AI method, AI only needs to measure 18 beam pairs hence reducing measurement time by 67%, expediting overall beam management process.
- Potential specification impact includes
 - Framework and procedures on exchange of neural network related information between network and UE that facilitate learning;
 - Exchange of neural network information for aligned L1/L3 measurement for inter-cell mobility procedure;
 - L1/L3 inter-cell mobility procedures optimized with the AI/ML enhancement;



Parameters: SLS, UMi 38.901, 7 cells, 3 sectors for each cell, UE speed 3km/h, carrier frequency 30GHz, subcarrier spacing 120KHz, 64 gNB antenna ([Mg Ng M N P] = [1 1 4 8 2]), 16 UE antenna ([Mg Ng M N P] = [1 1 2 4 2]). L1-RSRP based beam selection method. For each cell, 32 gNB beams and 16 UE beams, and then total 256 beam pairs. UE measures beams from 6 neighboring cells and then total 256*6=1536 beam pairs.

Example Case7: AI/ML+ Channel prediction

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



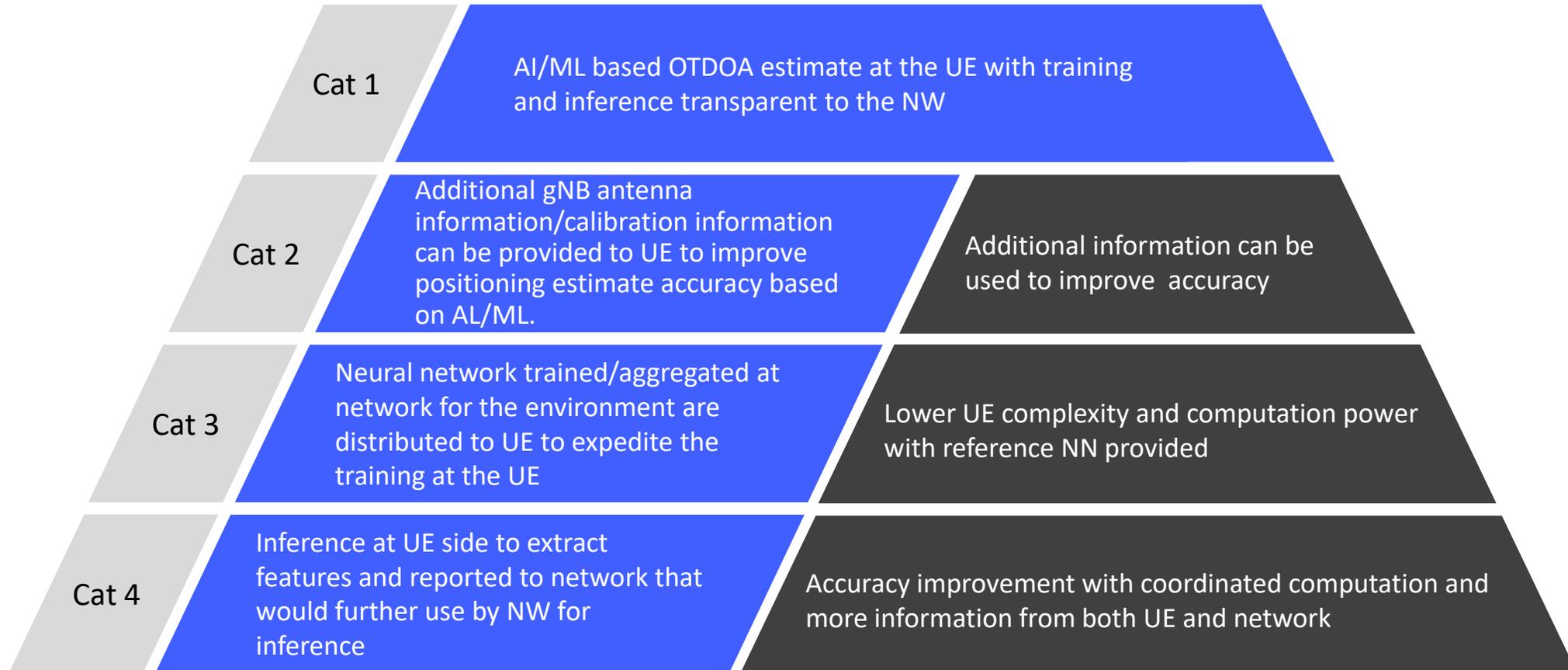
Case	NMSE	
	Speed 3km/h	Speed 30km/h
No channel prediction	0.15	1.6
AI with information of 2 periods	2.5e-8	2.6e-7
AI with information of 4 periods	3.3e-9	4.2e-8
AI with information of 6 periods	7.9e-10	1.9e-8
AI with information of 8 periods	5.0e-10	5.6e-9
AI with information of 10 periods	3.3e-10	4.2e-9

- Potential specification impact:
 - Framework and procedures on exchange of neural network related information between network and UE that facilitate learning and inference;
 - Exchange of neural network information for CSI report that facilitates channel/beam prediction;
 - Related measurement and reporting procedures;

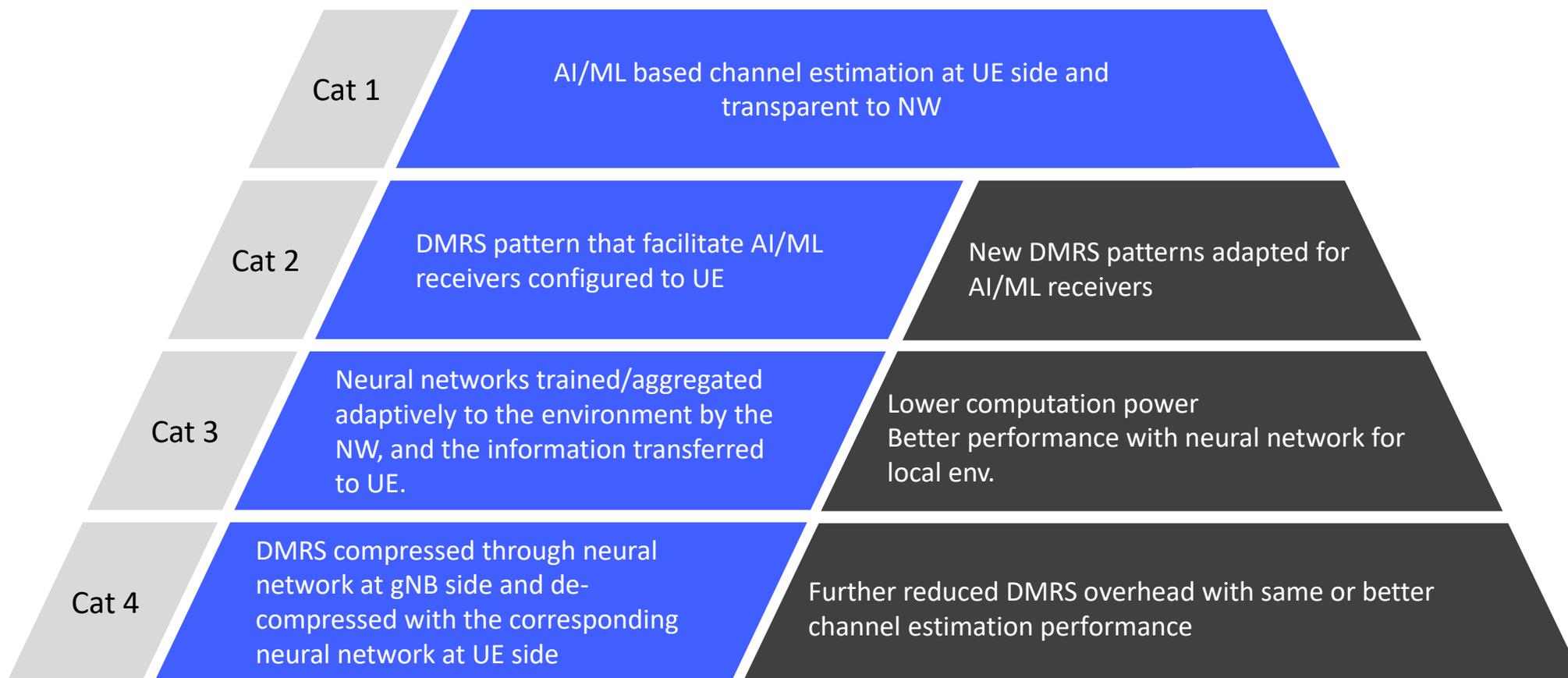
Categorization of use cases

- 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 assisted training 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



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



Channel Estimation

Potential use cases that would benefit from AI/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

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

Considerations for evaluation methodology

- 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;
 - Statistical distribution of local radio environment is typically fixed or slow changing;
 - Statistical model in 38.901 is representative for various practical radio environment;
 - 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 (including model structure and parameter numbers) ;

Potential SI/WI timeline and scope

- Initial phase (6~9 months): Potential objectives for RAN Plenary level study item:
 - Collection of AI/ML background information and terminology alignment;
 - Identify and categorize use cases that may benefit from AI/ML application on air interface based on for example:
 - Whether AI/ML application is transparent to the other side
 - Whether the exchanged information is directly related to training or inference
 - Recommendation of focused categories and use cases for the follow-up WG level study in Rel-18;
- WG study phase (9~12 months) : Potential objectives for RAN1-led Study item in Rel-18
 - Evaluation of performance of applying AI/ML in the identified areas, including detailed evaluation assumptions, metrics for comparison, whether and how to align DNN models between different companies, whether and how to construct data set for training/validation/test;
 - Framework and procedures of applying AI/ML in the identified areas
 - Potential specification impact of applying AI/ML in above areas, including related procedures;
- Work item phase to be started based on the output from the WG study item (Maybe started in Rel-19)

THANK YOU.

谢谢。