**3GPP TSG- RAN WG1 Meeting #110 R1-220xxxx**

**Toulouse, FR, Aug 22th – Aug 26th, 2022**

Agenda Item: 9.2.2.2

Source: Moderator (Apple)

Title: Email discussion on other aspects of AI/ML for CSI

# Introduction

This paper summarizes the discussion for agenda item 9.2.2.2.

# Finalize representative sub use cases for CSI feedback enhancement

## Summary of proposals

Spatial-frequency domain CSI compression using two-sided AI model is selected as one representative sub-use case in RAN1 109-e meeting. Study of other sub use cases is not precluded.

Following table summarizes company’s proposals on other use cases.

|  |  |
| --- | --- |
| **Company** | **Key Proposals/Observations/Positions** |
| Huawei | Proposal 1: For AI/ML-based CSI feedback enhancement, study temporal-spatial-frequency domain CSI compression using two-sided AI/ML model. |
| TCL communication | Proposal 1: The basic CSI feedback model based on auto-encoder reduces feedback bits through the air-interface, compared to the CSI feedback based on codebook. It is a functional replacement of the CSI feedback based on codebook.  Proposal 2: To fix the problem of outdated CSI feedback, the predictive CSI feedback model is necessary to predict CSI at scheduling time.  Proposal 3: Multiple CSI measurements can be compressed together and feedback at one shot to further reduces the feedback overhead.  Proposal 4: The CSI feedback compression along the frequency dimension, can be designed with new sparser CSI-RS patterns. The resource utilization is improved by allocating more REs to data transmission. |
| vivo | Proposal 1: To ensure the enhancement of CSI at both low and high-speed scenarios, study AI/ML for time domain CSI prediction with high priority. |
| ZTE | Proposal 3: For improving the CSI accuracy based on traditional codebook design using one-sided model, enhancement on Rel-16/17 eType II should be considered as a representative sub-use case for further study.  Proposal 4: For CSI prediction, we need to identify at least three aspects for further study: CSI prediction in spatial domain, frequency domain and time domain. However, time domain CSI prediction can be deprioritized in Rel-18 AI PHY, waiting for progress in Rel-18 MIMO session. |
| Sony | Proposal 1: RAN1 study paradigms for CSI measurement prediction to closer to the transmission resources allocated.  Proposal 2: RAN1 study paradigms for increasing CSI measurement granularity in both time and frequency for more accurate resource and MCS allocation.  Proposal 3: RAN1 study new methods of resource allocation and transport channel processing based on higher CSI measurement granularity.  Proposal 4: RAN1 should study how to reduce the transmission data size for both current coarse CSI and finer granularity CSI feedback. |
| Fujitsu | Proposal-1: AI/ML-based CSI feedback prediction in time domain should be selected as a sub-use case of CSI feedback enhancement.  Proposal-2: Both CSI prediction at UE side and CSI prediction at gNB side can be studied in this sub-use case. |
| Google | Proposal 7: Study CSI prediction as a use case for AI/ML based CSI feedback.  Proposal 8: AI/ML based joint CSI prediction and CSI compression should be deprioritized.  Proposal 9: AI/ML based CSI-RS configuration and overhead reduction should be deprioritized. |
| NEC | Proposal 1: Support temporal-domain CSI prediction using one-sided AI/ML model is selected as one representative sub use case.  Proposal 2: Support the adjustment of CSI feedback rate/ CSI reporting pattern based on the predicted CSI variation points as one of the final representative sub use cases. |
| Oppo | Proposal 1: The screening of CSI sub use cases needs to meet all the following conditions:  1. Potential performance gain.  2. Feasible evaluation methodology and valid training data set.  3. Reasonable non-AI/ML-based baseline for performance gain analysis.  4. Potential specification impacts.  Proposal 2: CSI compression(AE based Spatial-frequency domain CSI compression) should be given the high priority on evaluation and specification impact consideration. Other sub use cases, e.g. CSI prediction, CSI compression in temporal-spatial-frequency domain, or at gNB/UE side only could be given a lower priority until common understanding have been made |
| CATT | Proposal 1: The following sub use cases for CSI feedback are not considered in Rel-18:  – Temporal-spatial-frequency domain CSI compression;  – Improving the CSI accuracy based on traditional codebook design using one-sided model;  – AI/ML based DL/UL CSI prediction via UL/DL RS  – AI/ML based spatial/frequency/time domain CSI prediction through partial information;  – Resource allocation and scheduling;  – Joint CSI prediction and compression.  Proposal 2: The sub use case of AI/ML based CSI prediction in time domain is deprioritized in Rel-18. |
| Lenovo | Proposal 1 Agreements/Conclusions made in agenda 9.2.2.2 on further aspects of AI/ML for CSI feedback should focus on discussing the sub-use cases and corresponding specification impact. Details on the supported AI/ML model should be discussed in agenda 9.2.2.1 on evaluation of AI/ML for CSI feedback  Proposal 2 The study of spatial-frequency domain CSI compression using AI/ML should not be restricted by two-sided models in this stage. Decisions on the underlying AI model should be discussed in agenda 9.2.2.1 based on simulation and analytical results  Proposal 3 The study of temporal-spatial-frequency domain CSI compression is deprioritized  Proposal 4 CSI feedback overhead reduction and CSI accuracy improvement objectives are not to be treated in isolation, but into one sub use-case of CSI feedback enhancement  Proposal 5 Defer the discussion on AI-based CSI prediction sub use-case to RAN1#110bis-e  Proposal 6 CSI-RS configuration enhancement is not considered for study of AI/ML for CSI framework  Proposal 7 Resource allocation and scheduling sub-use case is discussed after the outline of CSI prediction sub-use case is finalized  Proposal 8 Use LTE UE-based sub-band selection for CQI reporting as a starting point for the study of AI-based resource allocation and scheduling  Proposal 9 Joint CSI prediction and compression is not considered |
| NVIDIA | Proposal 1: Autoencoder based CSI feedback enhancement should be selected as one representative sub use case.  Proposal 2: Focus on the sub-use case of spatial-frequency domain CSI compression using two-sided AI model to develop a thorough understanding of the performance of the AI model and the associated potential specification impacts.  Proposal 3: Potential sub-use cases other than of spatial-frequency domain CSI compression using two-sided AI model may be discussed at a later stage after a thorough understanding of the sub-use case of spatial-frequency domain CSI compression using two-sided AI model is developed. |
| Intel | Proposal 1:  • Spatial-frequency domain CSI compression using two-sided AI model should be prioritized |
| Xiaomi | Proposal 1: Other sub-use cases for CSI feedback are studied with lower priority except improving the CSI accuracy based on traditional codebook design using one-sided model due to its less specification impact. |
| CAICT | Proposal 1: CSI prediction and compression is treated as two separate sub use cases. |
| China Telecom | Proposal 1: Study benefits of using AI/ML for CSI compression in spatial and frequency domain compression using Rel-16 eType II codebook as a baseline.  Proposal 2: Study benefits of using AI/ML for CSI compression in Temporal-spatial-frequency domain compression.  Proposal 3: Further discuss AI/ML for CSI compression in Temporal-spatial-frequency domain compression as a possible sub-use case for CSI feedback enhancement.  Proposal 4: The performance gain of AI based CSI prediction may need FFS. |
| Samsung | Proposal 1-1: Study CSI prediction/extrapolation as one sub-use case for AI/ML for CSI feedback enhancement, including signaling requirements, input/output requirements, CSI configurations, and training strategies.  Proposal 1-2: Study CSI prediction/extrapolation at the UE under collaboration level (B), where limited information exchanges are required to configure/enable AI/ML.  Proposal 2-3: Study and verify model update of the encoder at the UE, where the gNB’s training strategy is not disclosed while transferring/configuring the AE.  Proposal 3-1: Study joint CSI prediction and compression as a representative sub-use case of AI/ML based CSI feedback enhancement.  • Consider joint CSI prediction and compression as temporal-spatial-frequency-domain compression. |
| LGE | Proposal #1: Prioritize to further study on legacy codebook enhancement and CSI-RS overhead reduction based on AI/ML. |
| Ericsson | Proposal 1 Add the temporal-spatial-frequency (TSF) domain compression as an optional variant of the two-sided model based spatial-frequency (SF) compression. The TSF variant allows multiple CSI-RS measurements over time to be utilized for CSI compression and prediction into the future. The proponent needs to explain whether SF or TSF was used when providing result and analysis.  Proposal 2 Study one-sided model-based CSI enhancements using traditional codebooks by investigating the possible benefits of UE to network codebook parameter recommendation and faster than RRC codebook re-configuration  Proposal 3 Study the performance and standardization impact of one-sided (i.e., UE) model based CSI prediction using the existing CSI framework as a starting point |
| CMCC | Proposal 1: Spatial-frequency domain CSI compression using two-sided AI model should be studied with high priority.  Proposal 2: Temporal-spatial-frequency domain CSI compression and CSI prediction could be studied with low priority. |
| Nokia | Proposal 3. Support CSI prediction as a second sub-use case.  Proposal 4. Compare channel prediction over broad bandwidth versus based on Type II CSI per sub-band.  Proposal 5: Consider UE sided as well as gNB sided channel prediction, as well as potentially include combined prediction between UE and gNB. |
| MediaTek | Proposal 1: Study CSI prediction under Release 18 AI/ML-based CSI enhancement. |
| Interdigital | Proposal 7: CSI prediction can be studied with lower priority if time allows in Rel-18 but there should be no conclusion/suggestion for CSI prediction as an outcome for normative work.  Proposal 8: Study CSI-RS configurations and overhead reduction at least to support CSI compression and prediction sub-use cases. |
| Qualcomm | Proposal 1: The study item should focus on the sub-use-case of spatial-frequency domain CSI compression and reconstruction of observed channel features using two-sided AI model; other sub-use-cases can be studied with lower priority. |
| Apple | Proposal 1: Consider time domain CSI prediction using one-sided AI model as one representative sub use case for R18 AI based CSI study. |
| AT&T | Proposal 1: Include resource allocation and scheduling as a sub-use case for CSI feedback enhancement.  Proposal 2: Include CSI prediction as a sub-use case for CSI feedback enhancements in combination with resource allocation and scheduling to further improve MU-MIMO scheduling optimization. |
| NTT DOCOMO | Proposal 1: Prioritize the discussion of spatial-frequency domain CSI compression from other sub-use case in 9.2.2. |

A summary of supporting companies for each sub-use case is listed in the table. (Note: Please update the table if proposals are captured wrong.)

|  |  |
| --- | --- |
|  | Support |
| temporal-spatial-frequency domain CSI compression using two-sided model | Huawei, TCL, China Telecom, Samsung, Ericsson |
| CSI prediction | TCL, vivo, Sony, Fujitsu, Google, NEC, CAICT, Samsung, Ericsson, Nokia, Apple, AT&T |
| CSI accuracy based on traditional codebook design | ZTE, Xiaomi, Ericsson |
| CSI-RS configuration and overhead reduction | TCL, Samsung, LGE, Interdigital |
| Resource allocation and scheduling | Sony, AT&T |
| Joint CSI prediction and compression | Samsung, Ericsson.  (Suggest to merge with Temporal-Spatial-Freq domain CSI compression). |

It has been mentioned by many companies that one-sided model is easier to implement, with less specification impact and likely to have a much shorter time to market than two-sided model. Therefore, it is proposed to study at least one sub use case with one-sided model for CSI feedback enhancement.

## ***Proposal 2-1:***

***Include at least one sub-use case with one-sided model under Release 18 AI/ML-based CSI enhancement use case.***

Please provide your view in the following two tables. 1st table please enter whether support or not support. The second table is for additional comments.

|  |  |
| --- | --- |
| Supporting companies | MediaTek, Ericsson, vivo, Fujitsu |
| Objecting companies | Lenovo, OPPO, CAICT, Futurewei, NVIDIA, NTT DOCOMO, Qualcomm, CATT |

|  |  |
| --- | --- |
| **Company** | **View** |
| Lenovo | We have concern with this kind of proposals that focus on the solution type rather than the use case. As per the RAN1 chair’s guidance, the objective of this agenda is to decide on sub-use cases to study corresponding to AI-based CSI feedback, and not to select a sub-use cases based on a given type of solution. We do not support this proposal |
| OPPO | For this proposal, we share the similar view with Lenovo.  And for the CSI prediction part, we think it could be a candidate sub use case for discussion. But before we agree that CSI prediction should be identified as an official sub use case, we think it is better to clarify some concerns first. For example, we need to consider how to deal with the following issues:  1. There should be clear and reasonable assumptions for simulation, especially the assumption on data set construction, e.g. do we make predictions based on the whole channel information (H domain)? If so, what is the bandwidth should be evaluated? And how many historical samples in the time domain would be used?  - for example, under the conditions of 32T4R and 52RB (624 SC), for the whole channel H, a time-domain sample is about 80K complex number. Considering multiple samples (e.g. 300K samples as many companies used in CSI compression), the overall training set size will be extremely large, about 200G byte (@ 32bit storage). The complexity for simulation needs to be considered.  - if we reduce the bandwidth or the antenna number for evaluation, whether the results on less RBs and antennas can be used to make the final conclusion for the TR needs to be considered as well.  2. Reasonable baseline, such as the comparison of CSI prediction by traditional methods and AI based CSI prediction |
| CAICT | We also have concerns on sub use case selection based on one-side or two-side model. |
| Futurewei | We share similar view as other companies and suggest deferring additional sub use cases till later. |
| MediaTek | We think it is good to study different type of model for CSI feedback enhancement. Even in BM, we are studying two sub use cases. Thus, we can study one additional sub use case. Among the candidates, we suggest studying CSI prediction which can provide overhead and latency reduction which are main goals of this use case. |
| NVIDIA | It’s better to focus on spatial-frequency domain CSI compression first. |
| Xiaomi | We have save view with Lenovo. We should focus on sub use case selection rather than AI/ML model Type. As described in our contribution, CSI accuracy based on traditional codebook design has less impact on specification, which can be considered to study as a representative sub use case. |
| ZTE | For the sub use cases of one-sided model, we should identify the pros and cons, and then clarify which sub use case needs to be studied. From our view, CSI accuracy based on traditional codebook design can be discussed since it closely complies with the existing protocols. However, CSI prediction can be deprioritized since simulation assumptions and reasonable baseline are not clear. |
| Samsung | Could you clarify if this sub-use case corresponds to CSI prediction? Is this sub-use case distinct from CSI compression? Some companies have proposed approaches to CSI compression that utilize one-sided models; we would prefer to study another sub-use case that is distinct from CSI compression, though. |
| NTT DOCOMO | We would rather not support this proposal, because the number of representative sub use case per scenario is not important. If one specific sub use case has promising gains, we should study that sub use case regardless of its framework difference from the other sub use case.  Also, as one-sided model will be studied even in BM and positioning, the similar mechanism can be reused after the discussion of BM and positioning if necessary. |
| Qualcomm | The specification impact for two-sided models is likely to be a superset of the one-sided model case. Additionally studying a one-sided model scenario may not add much more insight towards the “formulation of a framework” (as mentioned in the SID) for AI/ML-based air interface. Since a sub-use-case based on two-sided model is already agreed as a representative sub-use-case, we prefer the study item to focus efforts on this with higher priority. Besides, there are several sub-use-cases which use one-sided model. Instead of grouping them together, each should be treated individually in terms of how it functions and what are the benefits. |
| Ericsson | Although we support the proposal, we agree with Lenovos comment that an agreement should be targeting a sub-use case and not a solution type. We prefer to me more concrete e.g. “UE side CSI prediction” |
| LG | Agree with Samsung that the description of sub use case should be clear. |
| vivo | For CSI feedback enhancement, we think it is necessary to study more thoroughly by considering the impacts of both one-sided and two-sided model. CSI compression is a two-sided model. For one-sided model, we think time domain CSI prediction is a good choice. This is because the time domain CSI prediction can make up the performance loss at moderate and high-speed scenarios and make the picture of CSI enhancement complete (CSI compression is powerless to this performance loss at moderate and high-speed scenarios). To ensure the enhancement of CSI at both low and high-speed scenarios, we suggest to study AI/ML for time domain CSI prediction using one-sided model. |
| CMCC | We prefer to directly choose a certain sub use case, rather than based on one-side model or two-side model. |
| CATT | Similar view as the majority that we should select sub use cases with explicit definition, not based on where it resides.  We prefer to focus on spatial-frequency domain CSI compression at this stage. |
| Spreadtrum | We also prefer to discuss sub use case firstly. |
| Fujitsu | Considering the significant differences between one-sided model and two-sided model, we support to have discussion on potential sub-use case with one-sided model. Since AI/ML applying to CSI prediction is a promising sub-use case, we prefer to take it as the sub-use case, and study it using one-sided model. |

It has been mentioned that the potential specification impact of the temporal-spatial-frequency domain CSI compression using two-sided model is very similar to spatial-frequency domain CSI compression, and the temporal-spatial-frequency domain CSI compression can be an optional sub-use case for CSI compression using two-sided model. It would be beneficial to understand if any additional specification impact is identified to enable the temporal-spatial-frequency domain CSI compression sub-use case.

## ***Proposal 2-2:***

***Identify additional specification impact, if any, to support temporal-spatial-frequency domain CSI compression on top of spatial-frequency domain CSI compression using two-sided model.***

Please provide your view in the following two tables. 1st table please enter whether support or not support. The second table is for additional comments.

|  |  |
| --- | --- |
| Supporting companies | Huawei, Hisilicon, Google, Ericsson |
| Objecting companies | MediaTek, Qualcomm, vivo |

|  |  |
| --- | --- |
| **Company** | **View** |
| Lenovo | We are OK with studying temporal-spatial-frequency domain CSI compression as an extension to the already supported sub-use case for spatial-frequency domain compression. However, our preference is to defer study of this sub-use case until we agree on the outline for spatial-frequency domain compression, which can be the baseline for temporal-spatial-frequency domain CSI compression |
| OPPO | Better to focus on spatial-frequency domain CSI compression first. |
| CAICT | We would like to focus on spatial-frequency domain CSI compression first. |
| Futurewei | We are ok if some companies would like to study this sub use case but we suggest deferring additional sub use cases till later. |
| MediaTek | Instead of temporal-spatial-frequency domain CSI compression, we prefer to study CSI prediction which can provide different aspects of standardization impact. |
| Huawei, Hisilicon | In our simulation R1-2205890, temporal-spatial-frequency domain CSI compression using two-sided model can achieve a remarkable throughput gain on top of spatial-frequency domain CSI compression (additional 10%-15% THP gain), while it mostly reuse the similar input/output with little additional EVM. So discussing spec impact of temporal-spatial-frequency domain may not bring large additional efforts. |
| NVIDIA | It’s better to focus on spatial-frequency domain CSI compression first. |
| Xiaomi | We prefer to studying spatial-frequency domain CSI compression first. |
| ZTE | From our view, we may better defer the discussion of temporal-spatial-frequency domain CSI compression after the details of spatial-frequency domain CSI compression are settled, since the assumptions and baseline are still needed for further study. |
| Samsung | In our Tdoc, we mentioned several open issues for this sub-use case that could have spec impact, e.g.:   * mapping of feedback bits to the UCI payload * CSI-RS configurations (if necessary) * Input/output measurement combining, e.g., how past reports are combined to reconstruct the current report. |
| NTT DOCOMO | We prefer to prioritize the spatial-frequency domain CSI compression first. In the current RAN1 discussion, we have not identified the clear specification impacts brought by even spatial-frequency domain CSI compressions. After identifying them, we should discuss this proposal. |
| Qualcomm | We prefer to focus on spatial-frequency domain compression first. Temporal compression has not yet been agreed as a selected sub-use case; the proposal should be discussed after that. |
| Ericsson | As we commented in our tdoc, the RAN1 spec difference between TSF and SF is rather small (e.g. configuration of CSI-RS resources) and we can thus leave up to each company to optionally include TSF results in addition to their presented SF results. |
| LG | Whether T-S-F CSI compression is representative sub use case should be determined first. Also, regarding T-S-F CSI compression, there are two different interpretation, 1) w/o CSI predication 2) w/ CSI prediction (i.e., joint compression + predication). So, the description of this sub use case should be clarified. |
| vivo | Instead of temporal-spatial-frequency domain CSI compression, we prefer to study time domain CSI prediction which can provide different aspects of standardization impact. Furthermore, the enhancement by introducing the time domain CSI prediction is more valuable for moderate and high-speed scenarios (temporal-spatial-frequency domain CSI compression only is not targeted for these scenarios), which can complete the picture of CSI enhancement. |
| CMCC | We are ok with this sub-use case, but we prefer to study it until we have the outline of spatial-frequency domain CSI compression. |
| CATT | Prefer to focus on spatial-frequency domain CSI compression first. |
| Spreadtrum | Not support. We prefer to study CSI prediction in time domain, for it is beneficial for high/medium velocity UEs. |
| Fujitsu | Due to the heavy evaluation workload, we prefer to prioritize the sub-use case of spatial-frequency-domain CSI compression. |

# Potential specification impact for CSI compression with two-sided model

## Training collaboration

Following table summarize company’s proposals related to training collaboration.

|  |  |
| --- | --- |
| Huawei | Proposal 2: Study the potential specification impact for following training types:  • Type 1: On-network training with model transfer to UE  • Type 2: On-UE training with model transfer to network  • Type 3: Joint training across network and UE without model transfer, where the UE-side CSI generation part and the network-side CSI reconstruction part are trained in one forward propagation (FP) & backward propagation (BP) loop with necessary gradients exchange  • Type 4: Separate training at network and UE without model transfer, where the UE-side CSI generation part and the network-side CSI reconstruction part are trained by UE and network, respectively, in their own FP & BP loops |
| vivo | Proposal 6: Adopt either Definition Alt 1 or Definition Alt 2 to elaborate the process of joint training and separate training in CSI compression, although Definition Alt 2 is clearer with more detailed description.  • Clarification: Joint training could be done both at single NW node and multiple NW nodes (e.g., through gradient exchange between nodes).  Proposal 7: Study all 4 training options for CSI compression in short and long term, including performance, life cycle management, expected spec impacts, pros/cons, etc. |
| ZTE | Proposal 1: In CSI compression using two-sided model, the following mechanisms for AI/ML model training will be further studied:  ▪ Type 1: On-network training of two-sided model with model transfer to UE  ▪ Type 2: On-UE training of two-sided model with model transfer to NW  ▪ Type 3: Joint training of two-sided model without specified interactions between UE and NW  ▪ Type 4: Joint training of two-sided model with specified interactions between UE and NW  ▪ Type 5: Separate training at UE side and NW side for CSI feedback generation model / CSI reconstruction model respectively  Note: Joint training means the generation model and reconstruction model should be trained in the same loop for forward propagation and backward propagation. However, generation model and reconstruction model for separate training are trained in different loops for forward propagation and backward propagation. |
| Fujitsu | Proposal-3: For the two-sided AI/ML-based method, the study of STD impacts may start from:   Studying the CSI report configurations and the procedures related to the two-sided AI/ML model.   Studying the procedures related to pairing the AI/ML CSI generation part and the AI/ML CSI reconstruction part.   Studying the AI/ML ON/OFF switching and the AI/ML operation alignment for the two-sided AI/ML model. |
| Panasonic | Proposal 1: The following training collaboration should be studied.   Option 1: Joint training of the two-sided model with model transfer to UE.   Option 2: Joint training of the two-sided model with model transfer to network   Option 3: Joint training in offline engineering with multi-vendor agreements   Option 4: Separate offline training at UE side and network side for CSI feedback generation model / CSI reconstruction model respectively |
| Oppo | Proposal 3: A clear distinction between the training phase and the deployment phase would be helpful.  Training – The training procedure is to obtain a new AI/ML model. It starts from data collection and end up with an AI/ML model well trained and delivered to a given node, e.g. UE or gNB.  Deployment – The deployment phase refers to that after a node receives a model, some engineering operations are required to make the model available to use at that node, e.g. specific optimization, compiling and testing.  Proposal 5: In Rel-18, discussions on training collaboration should be handled with low priority.  - No collaboration needs in 3GPP could be treated as the basic assumption  - Collaboration needs for joint training with model transfer could be discussed with lower priority  - Collaboration needs for separate training with intermediate results interaction should be handled in subsequent releases. |
| Futurewei | Proposal 1: Study potential standards impact related to training collaboration Type 1, 2 and 4 for AI/ML model training.  Note: the assumption is that training collaboration Type 3 doesn’t incur specification impact in the model training phase. |
| CATT | Proposal 4: Study whether separate training for spatial-frequency domain CSI compression using two-sided AI/ML model is feasible, with the following aspects considered:  – Whether the same training dataset is used by both sides;  – Mechanisms on training dataset collection & transfer;  – Whether the two sides can use different AI/ML model structures (e.g. one side uses transformer and the other side uses ResNet);  – Whether associated /partial of AI/ML model related information exchange is needed. |
| Lenovo | Proposal 10 Study the advantages/disadvantages of network-based AI model training vs. UE-based AI model training |
| NVIDIA | Proposal 4: Study different training methods for autoencoder based CSI feedback including at least the following options:  • Option 1: Network performs autoencoder training  • Option 2: UE performs autoencoder training  • Option 3: Joint training between network and UE via federated learning  • Option 4: Joint training between network and UE via split learning |
| Xiaomi | Proposal 2: The training collaboration type of a two-sided model should be discussed with lower priority. |
| CAICT | Proposal 2: For CSI compression using two-sided AI model, offline training at gNB side with model transfer to UE side could be considered as baseline.  Proposal 3: Further discussion on standard impact of separate training and joint training for two-sided AI model is required. |
| China Telecom | Proposal 5: Further evaluation the performance impact of air-interface enhancement without model exchange. |
| Samsung | Proposal 2-2: Study CSI compression at the UE under collaboration level (D), where model exchanges are required to configure/enable AI, and inference is performed at the UE and the gNB. |
| Ericsson | Proposal 4 Aim to agree in RAN1#110 on the definitions of the four training collaboration types for CSI enhancements, a starting point is the draft shown above.  Proposal 5 For Type 3, study whether somewhat aligning AI/ML modes across different bilateral trainings (e.g. node internal model transfer approaches) with different vendors can reduce the implementation complexity due to the multi-vendor training situation.  Proposal 6 Study two-sided AI-based solutions for CSI reporting that enable UE side AI/ML / NW side AI/ML interoperability between different vendors, without the need for joint training. |
| CMCC | Proposal 5: In CSI compression using two-sided model use case, the Type 1 of offline AI/ML model training collaboration could be further studied:  • Type 1: Joint training of the two-sided model with model transfer to UE  Proposal 6: In CSI compression using two-sided model use case, transfer learning-based method could be studied for the training phase. |
| ETRI | Proposal 1: When considering Autoencoder for CSI feedback enhancement, first study the feasible training process of Autoencoder in both UE (Encoder) and gNB (Decoder) side supporting compatibility. |
| MediaTek | Proposal 2: Use the Alt 2 as the definition of joint training and separate training of two-sided model  Proposal 3: Discuss spec impact for offline/online training collaboration after the discussion on general UE-gNB collaboration is finalized in AI 9.2.1. |
| Qualcomm | Proposal 3: For the CSI feedback enhancement use-cases, focus on collaboration level “y” based on offline training that would enable a model to be optimized specifically for the device that will run the model.  Proposal 4: The distributed or separate offline training scenarios can be considered as a baseline for two-sided AI/ML model training. |
| Apple | Proposal 3: Consider four types of training collaboration for two-sided CSI compression use case.  • Type 1: NW trained and at least encoder is transferred to UE.  • Type 2: UE trained and at least decoder is transferred to NW  • Type 3: Offline engineering event where UE trained encoder/gNB train decoder through split learning  • Type 4: Separate training with training collaboration |

Training collaboration has been discussed in RAN1 109-e. It is desirable to agree on the training collaboration in RAN1 110. In 9.2.1 discussion, definition of “on-device training: online training at the UE” is discussed and not agreed yet. To avoid confusion, the proposals using “On-UE training” and “On-NW training” terminology is revised. The proposal reuse the wording before RAN1 109-e summary, with revision from proposals above.

Many companies share the view on prioritization of one or more types of training collaboration over the others. Views are not converging at this early phase. Therefore, we should have a detailed study of pros/cons of each approach.

## ***Proposal 3-1:***

***In CSI compression using two-sided model use case, the following AI/ML model training collaborations will be further studied:***

* ***Type 1: Joint training of the two-sided model with model transfer to UE.***
* ***Type 2: Joint training of the two-sided model with model transfer to NW.***
* ***Type 3: Joint training across network and UE without model transfer.***
* ***Type 4: Separate training at network and UE without model transfer, where the UE-side CSI generation part and the network-side CSI reconstruction part are trained by UE and network, respectively, in their own forward propagation and backward propagation loops.***
* ***FFS: Model fine tune.***
* ***FFS: Type 3 with specified interactions between UE and NW.***

***Note: Joint training means the generation model and reconstruction model should be trained in the same loop for forward propagation and backward propagation. Joint training could be done both at single NW node or across multiple nodes (e.g., through gradient exchange between nodes).***

Please provide your view below:

|  |  |
| --- | --- |
| **Company** | **View** |
| Lenovo | Support in general, propose the following rewording for clearer scope: ***At least for spatial-frequency ~~In~~ CSI compression using two-sided model use case, study potential specification impact corresponding to the following AI/ML model training collaborations ~~will be further studied~~:*** |
| OPPO | Support. Clear definitions for different training types and detailed analyses of pros/cons of each approach would be helpful for this study.  But we want to point out that it is not necessary to assume the subsequent discussions need to be done based on the analysis of different training types.  For different training types, actually the final result is specific models are obtained at specific nodes, waiting to be deployed, optimized and used. Therefore, subsequent discussions can be conducted based on this assumption. |
| CAICT | We support to have further combination of sub case description of CSI compression with collaboration level. The details could be further modified after the agreements on collaboration level and joint/separate training. |
| Futurewei | We are ok with the proposal. However, studying all 4 types may require a lot of effort. We suggest supporting Type 1 and Type 2 as baseline while Type 3 and Type 4 will not be excluded.  In addition, “model fine tuning” shouldn’t belong to one of the collaborations. |
| MediaTek | Support in principle. Definition of separate training is quite broad. To avoid confusion, it is better to use “UE-first training” or “NW-first training” as QC proposed. At least, we can add some note to clarify this scheme more like the note for joint training. |
| Huawei, Hisilicon | We are generally fine with the direction, but with two comments:  1, For type 1/2, as the terminology “on-UE training” may be confusing, maybe we can name it as “UE side”, otherwise it is not fully clear where the model is trained and who is to transfer the model. “Joint” is also removed, to distinguish with the “joint” under Type 3.  2, if there is a need to FFS the specified interactions between UE and NW, then both Type 3 and Type 4 should be included.  ***In CSI compression using two-sided model use case, the following AI/ML model training collaborations will be further studied:***   * ***Type 1: ~~Joint~~ training of the two-sided model at network side with model transfer to UE.*** * ***Type 2: ~~Joint~~ training of the two-sided model at UE side with model transfer to NW.*** * ***Type 3: Joint training across network and UE without model transfer.*** * ***Type 4: Separate training at network and UE without model transfer, where the UE-side CSI generation part and the network-side CSI reconstruction part are trained by UE and network, respectively, in their own forward propagation and backward propagation loops.*** * ***FFS: Model fine tune.*** * ***FFS: Type 3/4 with specified interactions between UE and NW.***   ***Note: Joint training means the generation model and reconstruction model should be trained in the same loop for forward propagation and backward propagation. Joint training could be done both at single NW node or across multiple nodes (e.g., through gradient exchange between nodes).*** |
| NVIDIA | Support this proposal, as the categorization would be useful further study. |
| Xiaomi | We support the proposal. The pros and cons can be further studied and discussed. But we should focus on the discussion on specification impact of AI/ML model deployment and inference first. |
| ZTE | We generally agree with the proposal. However, we should clarify whether these types are performed for offline training or online training. From our view,Type 1,2,4 can be conducted in both offline and online phase. Though Type 3 is for offline training, we think it can be reworded as **Joint/Separate training across network and UE without model transfer.** |
| Google | We do not know how Type3 and Type4 can work. We suggest adding FFS for Type3 and Type4. |
| Samsung | Before discussing this proposal, we should converge on definitions of “joint” and “separate” training. The FL provided Proposal 3.1.1.2-1 in the last meeting that included two alternatives for these definitions, so we should discuss that proposal first.  Once we converge on these definitions (either in this agenda item or in 9.2.1), then we can discuss this proposal. |
| NTT DOCOMO | Support the proposal. |
| Qualcomm | First, it would be important to clarify that all these types are for offline training involving training entities such as servers. The distinction should be made clear between (1) offline training between servers and (2) online training between gNB and individual UEs by exchanging gradients over-the-air. The current wording does not capture this distinction and could therefore lead to different interpretations.  Next, regarding Type 1 and Type 2, the UE-side and NW-side models are trained by a single entity. However, it is not clear why this should necessarily imply model transfer. As explained in our contribution, there could be a ‘centralized offline training’ scenario where a single entity trains both UE-side model and NW-side model and provides the models to the UE-side vendor and NW-side vendor respectively for device-specific compilation/optimization. Subsequently, the compiled model can be delivered to the UE / gNB over-the-top in a proprietary manner, i.e., without model transfer using the air-interface. We propose to add a new type to cover this case:  ***Type 5: Joint offline training of the two-sided model by a single entity without model transfer.***  As compared to Type 1 and 2, this approach is more feasible in the near term as it allows the UE-side and NW-side vendor to compile and optimize the model in a target-specific manner. |
| Ericsson | This is a good starting point, but I need to be clarified where the training approaches in Huawei and Qualcomm tdocs are categorized into these types? These ideas seems to be   * Sequential training starting with UE side training, or * Sequential training starting with NW side training   These could be seen as subcategories to Type 4 |
| vivo | We are generally in supportive of the proposal, and we also agree with the comments from Huawei. We believe that all 4 training collaboration options should be considered in this SI with no clear priority among them. From the perspective of specification impact, type 3 could serve as a baseline for type 1, 2 and 4, especially for some key LCM components such as data collection, model switch, performance monitoring, etc.  Our comment is that whether “model fine tune” should be discussed in training collaborations is unclear now. From our understanding, model fine tune could be done in all types. More clarifications are needed here. |
| CMCC | We are generally ok with this proposal. And maybe a clear definition of joint training and separate training could make this proposal clearer. |
| CATT | Prefer to discuss the proposal after definitions of “joint training” and “separate training” have reached consensus, or delete “joint training” in the first two types. One suggested change of the four types regardless the definitions of “joint training” and “separate training” is as follows:   * ***Type 1: ~~Joint training of~~Training the two-sided model at NW with model transfer to UE.*** * ***Type 2: ~~Joint training of~~Training the two-sided model at UE with model transfer to NW.*** * ***Type 3: Joint training across network and UE without model transfer, where the UE-side CSI generation part and the network-side CSI reconstruction part are trained at UE and network respectively, in the same loop for forward propagation and backward propagation.*** * ***Type 4: Separate training at network and UE without model transfer, where the UE-side CSI generation part and the network-side CSI reconstruction part are trained by UE and network, respectively, in their own forward propagation and backward propagation loops.*** |
| Spreadtrum | Support |
| Fujitsu | For Type 1 and Type 2, we may need to wait for the conclusion in 9.2.1 on deprioritizing model transfer related study.  In this sense, we think Type 3 and Type 4 can be studied first. Besides, model fine tune is very important to relax the request of generalization capability. Thus, we suggest to remove FFS for Model fine tune. |

## Data collection

Following table summarize company’s proposals related to data collection

|  |  |
| --- | --- |
| Huawei | Proposal 3: Study the potential specification impact on enabling network to obtain the ground-truth CSI from the realistic network as training labels for the AI/ML based model training at network. |
| Vivo | Proposal 2: In CSI compression, the following key LCM components for different training collaboration can be studied:  • Data collection  • Assistance information for model inference  • Model transfer  • Model activation/deactivation/switch  • Performance monitoring  • Model updating  • Assistance information for training  roposal 3: For data collection in CSI compression, study the potential specification impacts of:  • Data collection via current or potentially enhanced reference signals  • Data reporting via current or enhanced mechanism |
| Lenovo | Proposal 11 Study the means of feeding back the CSI training data from the UE to the network for FDD systems |
| NVIDIA | Proposal 5: For AI/ML model training for CSI feedback enhancement, study potential specification impact related to training data type/size, training data source determination, and assistance signalling and procedure for training data collection. |
| Xiaomi | Proposal 3: The signalling enhancements need to be studied for data collection used for training. |
| CMCC | Proposal 3: For AI based CSI enhancement, the potential spec impact on the training data reporting should be studied. |
| Qualcomm | Proposal 2: While generating the training dataset, the target CSI corresponding to a downlink measurement should be derived by the UE side.  Proposal 5: Study meta-data assistance signaling for UE’s training data collection for AI/ML Model development. Here, meta-data refers to auxiliary information about data such as an ID assigned for each distinct beam configuration. Meta-data can be used for scenario discovery during offline model development and scenario association during inference.  Proposal 6: UE data collection format and procedure is via offline engineering without need to involve 3GPP signaling.  Proposal 7: Focus on offline training scenario, where the development and training of the ML model for CSI feedback happens offline without the need to involve 3GPP signaling. |
| Apple | Proposal 2: Consider training assisted information in CSI-RS configuration for different training data set. |
| NEC | Proposal 3: Identify each training procedure. Especially, the exact procedures of separate training at UE and NW should be clarified to discuss the potential specification impacts. |

## ***Proposal 3-2:***

***In CSI compression using two-sided model use case, further discuss at least the following options for data collection:***

* ***Assistance signaling for UE’s training data collection for AI/ML Model development.***
* ***Transmission/reception of the datasets for training/validation/testing.***

Please provide your view below:

|  |  |
| --- | --- |
| **Company** | **View** |
| Lenovo | Support in general, propose the following rewording for clearer scope: ***At least for spatial-frequency ~~In~~ CSI compression using two-sided model use case, further discuss potential specification impact corresponding to at least the following options for data collection*** |
| OPPO | Support in general. We think the label data collection should be considered as well and propose the following rewording:  *In CSI compression using two-sided model use case, further discuss at least the following options for data collection:*   * *UE’s training data collection for AI/ML Model development, including data and assistance signaling* * *gNB’s label data collection for AI/ML Model development, including data and assistance signaling* * *Transmission/reception of the datasets for training/validation/testing.* |
| CAICT | Support. |
| Futurewei | We are ok with the proposal. |
| MediaTek | We generally support the proposal. We think it is better to downselect the collaboration types in proposal 3-1 instead of studying data collection for all different types. |
| Huawei, Hisilicon | Firstly, the data collection by network should also be studied as pointed out by OPPO. On top of that, we think the “for AI/ML Model development” is not clear, since the data collection is for training rather than model development. So “for AI/ML model development” is removed.  In addition, the last bullet of “Transmission/reception of the datasets for training/validation/testing” is not fully clear to us: are there some more clarifications?  Secondly, some assistant information would be proprietary and not exposed to the other side. We add a “note” borrowed from the conclusion of the 9.2.3.2 in the 109-e meeting.  ***In CSI compression using two-sided model use case, further discuss at least the following options for data collection:***   * ***UE’s training data collection ~~for AI/ML Model development~~, including data and assistance signaling if any*** * ***gNB’s label data collection ~~for AI/ML Model development~~, including data and assistance signaling if any*** * ***[Transmission/reception of the datasets for training/validation/testing.]*** * ***Note: The provision of assistance information may be infeasible due to the concern of disclosing proprietary information to the other side.*** |
| NVIDIA | Support to further discuss these aspects. |
| Xiaomi | We support the proposal in general. The collected data can be applied to AI/ML model training, validation, testing, fine-tuning. Therefore, the proposal can be reworded as:  ***In CSI compression using two-sided model use case, further discuss at least the following options for data collection for AI/ML model training/validation/testing/fine-tuning:***   * ***Assistance signaling for UE’s training data collection ~~for AI/ML Model development~~.***   ***Transmission/reception of the datasets ~~for training/validation/testing~~.*** |
| ZTE | We are not clear what model development means and suggest replacing it with Model training/inference. So we provide the rewording as follows:  *In CSI compression using two-sided model use case, further discuss at least the following options for data collection:*   * *UE’s data collection for AI/ML Model training/inference, including data and assistance signaling* * *gNB’s data collection for AI/ML Model training/inference, including data and assistance signaling* * *Transmission/reception of the datasets for training/validation/testing.* |
| Google | Is it for online training? |
| Samsung | We support further discussion of these options.  Transmission of datasets for training/validation/testing may have significant overhead. |
| NTT DOCOMO | Generally fine with the proposal. Assistance signaling for training is unclear. We think assistance signaling even includes the second bullet (Transmission/reception of the datasets for training/validation/testing) as well. To clarify the difference between the first and second sub bullets, we prefer to update “assistance information signaling” instead of “assistance signaling”. |
| Qualcomm | We agree with the first item – assistance signaling related to training data collection. Regarding the second point, in our view, the exchange of datasets could be done in a proprietary way for the purpose of offline training. The need for RAN1 specification impact for this purpose is not clear. |
| Ericsson | Can be further discussed after data collection has progressed in the general agenda 9.2.1. It is unclear at this point whether there are any specific issues related to CSI use case. Also, some of the work is likely RAN2 issue. |
| LG | Support with Lenovo’s clarification. |
| vivo | Generally fine with the proposal, but we have similar comments on the term “model deployment” as raised by Huawei and Xiaomi. Given that whether model deployment will be included in LCM procedure is under discussion in general framework, we suggest to temporally remove it until more clear definitions are drawn.  We are also fine with the rewording of UE’s training data collection and gNB’s label data collection. |
| CMCC | We are not clear with the definition of “model development”.  ZTE’s version is fine for us. |
| CATT | Is this proposal for training only or for both training and inference? We think data collection for model training and inference should be discussed separately. Since the inference related spec impact would be discussed in section 3.3, it is our view that this proposal is on data collection for training.  We suggest to change the wordings as follows:  ***In CSI compression using two-sided model use case, further discuss at least the following aspects for data collection:***   * ***UE’s data collection for AI/ML Model training, including data and assistance signaling, if any;*** * ***Network’s data collection for AI/ML Model training, including data and assistance signaling, if any;*** * ***Transmission/reception of the datasets for training/validation/testing.*** |
| Spreadtrum | In our understanding, the first sub-bullet can be covered by the second bullet. |
| Fujitsu | Support. |

## Inference related spec impact

Following table summarize company’s proposals related to inferencing

|  |  |
| --- | --- |
| Huawei | Proposal 4: Study the potential specification impact for configuration and content of AI/ML model input/output for CSI compression.  Proposal 5: Study the potential specification impact for the alignment of pre-processing and post-processing approaches between network and UE.  Proposal 6: Study the potential specification impact for the quantization/dequantization method for the compressed CSI. |
| Vivo | Proposal 8: The specification impacts of encoder (decoder) input depend on the corresponding training options:  • For training collaboration Type 1, 2, and 4, encoder input and decoder input should be specified  • For training collaboration Type 3, encoder input could be left to implementation and decoder input should be specified  Proposal 9: The specification impacts of encoder (decoder) output depend on the corresponding training options:  • For training collaboration Type 1, 2, and 4, encoder output and decoder output should be specified.  • For training collaboration Type 3, encoder output should be specified and decoder output could be left to implementation.  Proposal 10: Study the potential modifications in CQI/RI for AI/ML based CSI compression.  Proposal 14: Study the specification impact of pre- and post-processing approaches for model input/output to address the issue of adaption to multi-configurations (e.g., bandwidth, number of ports, feedback payloads, antenna configurations, etc.). |
| ZTE | Proposal 2: For spatial-frequency domain CSI compression using two-sided AI model, at least further study and evaluate the following options for different types of AI/ML model input/output,  ▪ Option 1: The input of AI encoder is a raw channel (i.e obtained directly from CSI-RS) without any further pre-processing and corresponding output is a recovered raw channel:   Option 1a: The raw channel is in frequency domain   Option 1b: The raw channel is in time domain  ▪ Option 2: The input of AI encoder is a precoding matrix which is pre-processed from a raw channel and corresponding output is a recovered precoding matrix:   Option 2a: The precoding matrix is a group of eigenvectors   Option 2b: The precoding matrix is an eType II-like PMI. |
| Panasonic | Proposal 2: For each option of training collaboration, configuration and content for CSI report should be studied.  Proposal 3: For each option of training collaboration, handling of rank of AI/ML model should studied. |
| Google | Proposal 2: Study the input of CSI compression based on the eigenvectors of the raw channel with a wideband precoder selected as SD basis, e.g. HW1. |
| Oppo | Proposal 4: The research and potential protocol impact analysis on the training, deployment, inference and management of AI/ML solutions could be discussed separately.  - The requirements of AI/ML solutions for standardization in the inference phase should be evaluated with high priority  - Corresponding necessary deployment and management issues also need to be analyzed  Proposal 10: Consider the impact of standards on encoder input, decoder output, as well as pre/post-processing(if needed) together.  Proposal 11: Consider the impact of standards on encoder output and decoder input together.  Proposal 15: Discussions on Quantization/Loss function should be left to implementation. |
| FutureWei | Proposal 2: Study potential standards impact regarding exchanging/sharing the main encoder input type, e.g., raw CSI feedback or eigenvectors while the exact input format, shape, and other accessory input features may be considered as implementation dependent.  Proposal 3: Study potential standards impact related to the new CSI feedback to be transmitted over the air-interface (i.e., in bits) that is generated from the encoder at UE side and may be used as the input to the decoder directly or further processed to generate the input to the decoder at gNB.  Proposal 4: Study potential standards impact related to other supporting information that is needed for the gNB to recover the unquantized encoder output from the received bits, e.g., quantization codebook, if such information is not already available.  Proposal 8: Study potential standards impact to support AI/ML model inference procedure at UE side and gNB side. |
| Lenovo | Proposal 12 Study CSI reporting configuration enhancement for AI-based CSI feedback under different network-UE collaboration levels  Proposal 13 Study CSI reporting content enhancement for AI-based CSI feedback under different network-UE collaboration levels  Proposal 14 Study means of signaling the AI model parameters for two-sided AI models |
| NVIDIA | Proposal 8: For AI/ML based CSI feedback, study potential specification impact related to report/feedback of model input for inference, type of model input, and model input acquisition and pre-processing.  Proposal 9: For AI/ML based CSI feedback, study potential specification impact related to report/feedback of model inference output and post-processing. |
| Intel | Proposal 2:  • Specification impact for the CSI compression using two-sided AI model may include the following  o Introduction of NN concept in specification for PMI bits calculation (encoder part)  o Introduction of NN concept in specification for precoder reconstruction for CQI (decoder part)  o RRC configuration (including NN parameters configuration and NN coefficients)  o UCI design (including new CSI quantity)  o UE capabilities (including CPU definition for the AI/ML CSI) |
| Spreadtrum Communications | Proposal 2: AI/ML module at UE side can be delivered from gNB, based on UE’s capability. The input format and output format of the AI/ML module also should be included.  Proposal 3: Aperiodic CSI reporting should be considered firstly.  Proposal 4: The configuration of CSI-ResourceConfig and/or CSI-ReportConfig should be enhanced. |
| Xiaomi | Proposal 5: Input/output data format of AI/ML model needs to be defined.  Proposal 7: In order to avoid performance loss due to mismatch between precoder and RI/CQI, it needs to study how to calculate RI/CQI when decoder is not deployed at UE side. |
| Samsung | Proposal 2-1: For CSI compression, study signaling requirements, input/output requirements, CSI configurations, and training strategies. |
| LGE | Proposal #2: Consider enhancement of CSI-RS/CSI reporting configurations for AI/ML based CSI feedback. |
| Ericsson | Proposal 7 Study model-based MIMO channel feature extraction methods (pre-processing) based on spatial- and frequency-domain DFT codebooks (using Type-II CSI Rel.16) and associated specification impacts (e.g., additional required signaling over the air interface).  Proposal 8 Study MIMO channel normalization methods, and associated specification impacts (e.g., additional required signaling over the air interface).  Proposal 9 Include Ais based on both real- and complex-valued NNs in the study, where each proponent report which type was used in evaluaitons (i.e., do not restrict the study to only to real-valued NNs).  Proposal 10 Study CSI enhanced reporting options for two-sided AI based solutions. For example, the CSI report may include a preferred rank indication, channel quality information, interference information, feature extraction information, and compression quality indicators.  Proposal 11 Study quantization methods for UCI, including quantization aware training and complex-valued activation functions. |
| Nokia | Proposal 1. Study channel-, eigenvector-based and W2-based two-sided models for CSI feedback compression.  Proposal 2. For the two-sided models, study the impact of quantizers on CSI feedback compression. |
| MediaTek | Proposal 4: Discuss spec impact for model input (encoder/decoder input) and pre-processing after the discussion on training collaboration is finished.  Proposal 5: Discuss spec impact for model output (encoder/decoder output) after the discussion on training collaboration is finished.  Proposal 8: Study potential spec impact on quantization for CSI compression with auto-encoder focusing on the followings  • Uniform vs Non-uniform quantization  • Scalar vs Vector quantization  • Derivable (approximated) quantization  • Gradient passing  • Learnable quantization offset |
| Interdigital | Proposal 5: Study the use of pre-processing in the frequency, spatial and angle-delay domains as means to reduce the AI/ML model complexity.  Proposal 6: Study specification impacts of CSI compression using AI/ML including: pre-processing the measurements at the UE, AI/ML model selection at the UE, new CSI report types and new CSI reporting mechanism. |
| Qualcomm | Proposal 8: The input to the UE-side model should be left to UE implementation, the output at the NW-side model can be specified  Proposal 9: Preprocessing at UE side is upto UE-implementation and should not be specified.  Proposal 10: For AI-based CSI feedback, the size of the UCI payload and the final CSI format can be specified. Post-processing of CSI decoder output to the final CSI should be specified as part of the final CSI structure.  Proposal 11: The discussion on specification impact on input/output of CSI encoder/decoder and CSI report configuration is independent of training collaborations. |
| Apple | Proposal 4: Input to the AI encoder including potential pre-processing needs to be signalled.  Proposal 5: Output of the AI encoder needs to be signalled, including RI, CQI, inferencing output and potential encoder neural network ID.  Proposal 6: RAN1 further discuss CQI definition for AI based CSI compression.  Proposal 7: Output of the AI decoder including potential post-processing needs to be signalled. |
| NTT DOCOMO | Proposal 2: Study the potential specification impacts based on CSI type for input/output. |

## ***Proposal 3-3-1:***

***In CSI compression using two-sided model use case, further study potential specification impact on CSI report, including at least***

* ***Encoder output and decoder input, including size/configuration and potential post/pre-processing of encoder output/decoder input.***
* ***RI***
* ***CQI with potential modification for AI based CSI compression.***

Please provide your view below:

|  |  |
| --- | --- |
| **Company** | **View** |
| Lenovo | Support in general. Regarding CSI reporting, we prefer considering RI, PMI and CQI for AI-based CSI enhancements. Therefore, we propose the following rewording for clarity  ***Proposal 3-3-1:***  ***At least for spatial-frequency~~In~~ CSI compression using two-sided model use case, further study potential specification impact on CSI reporting, including at least***   * ***Encoder output and decoder input, including size/configuration and potential post/pre-processing of encoder output/decoder input.*** * ***Consider at least the following report quantities for AI-based CSI feedback enhancement: RI, PMI, CQI*** * ***~~CQI with potential modification for AI based CSI compression.~~*** |
| OPPO | Support in general. Prefer leaving the RI and CQI parts in FFS |
| CAICT | Support |
| Futurewei | Exact input to the decoder is internal to gNB, which may directly come from dequantized bits received, or after further processing, which is up to implementation and should not have specification impact. It should be clarified why such information needs to be discussed / specified. |
| MediaTek | Support in principle. Can be discussed with quantization/dequantization |
| Huawei, Hisilicon | For encoder/decoder, change it to “CSI generation part”/”CSI reconstruction part”, respectively, to align the terminology.   * ***~~Encoder~~ CSI generation part output and ~~decoder~~ CSI reconstruction part input, including size/configuration and potential post/pre-processing ~~of encoder~~ for CSI generation part output/~~decoder~~ CSI reconstruction part input.***   Why RI is mentioned alone without CQI/PMI? |
| NVIDIA | Support to further study these aspects. |
| Xiaomi | Different input/output type or size of AI/ML model has impact on system performance. Hence, the first bullet can be discussed in evaluation of CSI feedback, i.e., agenda 9.2.2.1.  For CSI reporting quantities, we share similar view with Lenovo and agree with the rewording. |
| ZTE | Agree with the rewording of Lenovo. |
| Google | Support in principle |
| Samsung | Support |
| NTT DOCOMO | Support the proposal. In our view, if offline training is used and multi-vendor agreements are assumed before model deployments, the input/output and pre-processing/post-processing can be up to implementation. |
| Qualcomm | Support, except that the pre-processing / post-processing of encoder output/decoder input need not be specified. The methods used could be aligned between the two sides in an industry agreement as part of the offline training process. |
| Ericsson | It is not clear at this point whether RI and CQI should be included in the report, i.e. the study should consider both explicit and implicit (precoder, RI,CQI recommendation) type of CSI reporting. Pre and pro-processing needs perhaps to be specified so that gNB know what the received CSI represents. |
| LG | Support in principle, and we also not sure on RI. So, we can put FFS on RI. |
| vivo | We are in supportive of the proposal. |
| CMCC | Support in principle and prefer Lenovo’s version. |
| CATT | Support in principle.  In our understanding, the output of encoder and the input of decoder would be different or the same, which depends on whether quantization is included in the AI/ML model or not. |
| Spreadtrum | Support in principle |
| Fujitsu | **In last meeting it was agreed that:**  For the evaluation of the AI/ML based CSI compression sub use cases, a two-sided model is considered as a starting point, including an AI/ML-based CSI generation part to generate the CSI feedback information and an AI/ML-based CSI reconstruction part which is used to reconstruct the CSI from the received CSI feedback information.  Thus, we’d better use the wording CSI generation part and CSI reconstruction part rather than encoder and decoder.  **We support the first sub-bullet, but think RI and CQI should be further clarified or for further study.** |
| Sony | Support in principle |

## ***Proposal 3-3-2:***

***In CSI compression using two-sided model use case, further study potential specification impact on encoder input, including at least***

* ***Model input type/dimension/configuration***
* ***Model-based MIMO channel feature extraction methods (pre-processing) based on spatial- and frequency-domain DFT codebooks (using Type-II CSI Rel.16) and associated specification impacts (e.g., additional required signaling over the air interface).***
* ***MIMO channel normalization methods, and associated specification impacts (e.g., additional required signaling over the air interface).***
* ***Note: The encoder input can be left for implementation at least for training collaboration type 3.***

Please provide your view below:

|  |  |
| --- | --- |
| **Company** | **View** |
| Lenovo | Deprioritize until more clarity on AI-based CSI feedback design from agenda 9.2.2.1 is available |
| OPPO | Support in general, and share the similar view with Lenovo that the specific impact is close related to the discussion in Section 9.2.2.1.  For the proposal, prefer modifying the wording as:  *In CSI compression using two-sided model use case, further study potential specification impact on encoder input, including at least*   * *Model input type/dimension/configuration and potential post processing, e.g.* * *Model-based MIMO channel feature extraction methods (pre-processing) based on spatial- and frequency-domain DFT codebooks (using Type-II CSI Rel.16) and associated specification impacts (e.g., additional required signaling over the air interface).* * *MIMO channel normalization methods, and associated specification impacts (e.g., additional required signaling over the air interface).*   *Note: The encoder input can be left for implementation at least for training collaboration type 3.* |
| CAICT | Support in principle and fine to have further discussions after some conclusions from 9.2.2.1. |
| Futurewei | We think studying the impact on encoder input type is ok, while the exact format/dimension should be left to implementation and no need to be specified. |
| MediaTek | Fine with the proposal. |
| NVIDIA | These detailed aspects may be better addressed after more progresses are made under 9.2.2.1. |
| Xiaomi | This proposal can be discussed after some conclusions from agenda 9.2.2.1. |
| ZTE | We generally agree with this proposal, however, we think this issue is closely related to the sub use cases or training collaboration levels, which can be further discussed after those issues are stable. |
| Google | Support. |
| Samsung | Support |
| NTT DOCOMO | Fine with the proposal. However, we do not think there is specification impacts for input/pre-processing/post-processing without 3gpp-based model transfer.  In our view. pre-processing/post-processing also can be left for implementation, because these pre-processing/post-processing can be viewed as one implementation specific steps of AI/ML model. It should be treated as proprietary assets as well as the model structure. |
| Qualcomm | The need for specification impact of the encoder input is not clear. For proprietary models, the input should not be specified. For non-proprietary models also, the UE-side vendor would be involved in testing the model to confirm that the UE will be able to run it. Hence, there is no need for RAN1 signaling to inform the UE about the input format as this can be done in a proprietary manner.  In legacy CSI framework, the input and method used to derive the PMI is left to UE implementation. This allows the UE to optimize performance-complexity tradeoff in an implementation-specific manner. This principle should be preserved for ML-based CSI feedback also. Specifying the encoder input would go against this principle. Regarding the MIMO channel pre-processing, it would be sufficient to specify how the gNB would interpret the output of the NW-side model and how it would derive the output CSI. Specifying the input of the UE-side model does not add anything.  Also, it would be good to clarify the scope of each training type. Types 1 and 2 combine centralized training with model transfer. As discussed in our earlier comment, centralized training could also be done without model transfer over the air-interface. The UE-side vendor would compile and test the trained model and then give it to the UE in a proprietary manner. In such cases, the encoder input format can also be conveyed in a proprietary manner and need not be specified and signaled over the air-interface. There is no need to define RAN1 signaling to specify the input in this case. |
| Ericson | Support the proposal |
| LG | Support |
| vivo | We are in supportive of the proposal. |
| CMCC | We support in principle, but prefer discuss it later after some conclusions from agenda 9.2.2.1. |
| CATT | Support in principle and fine to have further discussions after there is some conclusions from 9.2.2.1. |
| Spreadtrum | Support |
| Fujitsu | ~~encoder~~ -> CSI generation part  For the second and third sub-bullet, we only need to say, pre-processing associated specification impacts, and do not need to restrict the study to such specific method.   * *Model input type/dimension/configuration* * *~~Model-based MIMO channel feature extraction methods (pre-processing) based on spatial- and frequency-domain DFT codebooks (using Type-II CSI Rel.16) and associated specification impacts (e.g., additional required signaling over the air interface).~~* * *~~MIMO channel normalization methods, and associated specification impacts (e.g., additional required signaling over the air interface).~~* * Pre-processing associated specification impacts |

## ***Proposal 3-3-3:***

***In CSI compression using two-sided model use case, further study potential specification impact on decoder output, including at least***

* ***Model output type/dimension/configuration and potential post processing***

Please provide your view below:

|  |  |
| --- | --- |
| **Company** | **View** |
| Lenovo | Deprioritize until more clarity on AI-based CSI feedback design from agenda 9.2.2.1 is available |
| OPPO | Support in general, and share the similar view with Lenovo that the specific impact is close related to the discussion in Section 9.2.2.1. |
| CAICT | Same view as Lenovo and OPPO. |
| Futurewei | We are ok with this proposal in general but we think the exact post processing may be left for implementation while the needs/impacts relevant to standards should be specified. |
| MediaTek | Support, but fine with deferring discussion later |
| NVIDIA | These detailed aspects may be better addressed after more progresses are made under 9.2.2.1. |
| Xiaomi | This proposal can be discussed after some conclusions from agenda 9.2.2.1. |
| ZTE | We generally agree with this proposal, however, we think this issue is closely related to the sub use cases or training collaboration levels, which can be further discussed after those issues are stable. |
| Google | Support |
| Samsung | Support |
| NTT DOCOMO | Fine with the proposal. In our view, if offline training is used and multi-vendor agreements are assumed before model deployments, the output format and post-processing can be up to implementation. |
| Qualcomm | Support with a suggested change:  *“In CSI compression using two-sided model use case, further study potential specification impact on output CSI ~~decoder output~~, including at least…"*  The output CSI is the final outcome after applying post-processing to the decoder output, and that should be specified. |
| Ericsson | Support. In principle the gNB need to know what the decoder output represents (full H channel? Eigenvectors? Etc..), whether this requires specifications needs to be studied. |
| LG | Support |
| vivo | Fine with the proposal. |
| CMCC | We support in principle, but prefer discuss it later after some conclusions from agenda 9.2.2.1. |
| CATT | Support in principle and fine to have further discussions after there is some conclusions from 9.2.2.1. |
| Spreadtrum | Support in principle |
| Fujitsu | *~~decoder~~->* CSI reconstruction part  We support this proposal in general. |

## ***Proposal 3-3-4:***

***In CSI compression using two-sided model use case, further study potential specification impact on the quantization/dequantization method for the compressed CSI.***

Please provide your view below:

|  |  |
| --- | --- |
| **Company** | **View** |
| Lenovo | Deprioritize until more clarity on AI-based CSI feedback design from agenda 9.2.2.1 is available |
| OPPO | Not necessary. Quantization should be left to implementation |
| CAICT | Not necessary. |
| Futurewei | We are ok with the proposal in general. However, how quantization is implemented should be left to vendors, but the corresponding standards impact part can be studied and discussed, e.g., quantization codebook to be exchanged between UE and gNB. |
| MediaTek | Fine to study but it can be discussed with proposal 3-3-1. |
| NVIDIA | These detailed aspects may be better addressed after more progresses are made under 9.2.2.1. |
| Xiaomi | This proposal can be discussed after some conclusions from agenda 9.2.2.1. |
| ZTE | From our view, various quantization/dequantization methods are implemented by companies and the related spec impacts can be further studied after quantization/dequantization discussion in agenda 9.2.2.1. |
| Google | Support |
| Samsung | Support |
| Qualcomm | The quantization and dequantization methods do not need to be specified. During the offline training process, the methods used could be aligned between the two sides in an industry agreement. |
| Ericsson | Support. The gNB may need to know how to interpret the bits in the received UCI to be able to provide the correct input to the decoder. |
| LG | Support |
| vivo | Fine with the proposal. |
| CMCC | We think the quantization/dequantization method is more related to implementation and should be discussed in agenda 9.2.2.1 for evaluation. |
| CATT | Similar view as MediaTek that it can be discussed together with (or be part of) proposal 3-3-1. |
| Fujitsu | Agree. |
| Sony | Support |

## Performance monitoring, model update, activation/de-activation/switching

Following table summarize company’s proposals related to model performance monitoring, activation/de-activation/switching.

|  |  |
| --- | --- |
| Huawei | Proposal 7: Study the potential specification impact for life cycle management for AI/ML-based CSI feedback, including dataset collection, model monitoring, model switching, and model updating.  Proposal 8: Study the potential specification impact for the co-existence between AI/ML mode and legacy non-AI/ML mode. |
| Vivo | Proposal 4: For model activation/deactivation/switch in CSI compression, study the potential specification impacts of following options:  • Model activation/deactivation/switch via model ID  Proposal 5: For performance monitoring in CSI compression, study the potential specification impacts of following cases:  • Complete model (encoder and decoder) is available at UE  • Complete model (encoder and decoder) is available at gNB  • Neither UE nor gNB has complete model  Proposal 13: Study the schemes and corresponding specification to address issues of scenario (e.g., Umi, UMa, Indoor, etc.) and configuration adaption (e.g., bandwidth, number of ports, feedback payloads, antenna configurations, etc.) in AI/ML for CSI feedback enhancement, including:  • Schemes that can be easily adapted to different scenarios and configurations  • Procedure and ignaling for scenario- and configuration-specific data collection;  • Procedure and ignaling for model selection;  • Signalling to indicate the application scope of each model |
| ZTE | Proposal 5: During study phase, companies need to evaluate and identify the solutions to perform model life cycle management in CSI feedback enhancement, at least following perspectives can be further studied:  ▪ Case 1: A common AI/ML model that is applicable to all scenarios  ▪ Case 2: AI/ML model switching to adapt different scenarios based on model performance monitoring  ▪ Case 3: An offline trained AI/ML model to be updated online |
| Fujitsu | Proposal-3: For the two-sided AI/ML-based method, the study of STD impacts may start from:   Studying the CSI report configurations and the procedures related to the two-sided AI/ML model.   Studying the procedures related to pairing the AI/ML CSI generation part and the AI/ML CSI reconstruction part.   Studying the AI/ML ON/OFF switching and the AI/ML operation alignment for the two-sided AI/ML model. |
| Panasonic | Proposal 4: The following options should be studied for life cycle management.   Solution 1: gNB side performance monitoring   1-1: UE transmit encoder input as CSI report periodically or occasionally.   1-2: gNB may directly use system throughput or ratio of NACK.   Solution 2: UE side performance monitoring   2-1: UE calculate decoder output using virtual decoder in UE.   2-2: UE may obtain the inference results indicated from gNB periodically or occasionally   2-3: UE may use PDSCH decoding performance as KPI. |
| Google | Proposal 6: Study the AI/ML model adaptation for CSI compression, where different AI/ML models may be with different compression ratio. |
| NEC | Proposal 3: Study the mechanism of evaluating model performance to facilitate AI/ML model lifecycle management.  Proposal 4: Study the behaviors of UE or (and) gNB after AI/ML model performance deteriorates, e.g., model update (fine-tuning), model switching, fall back to non-AI/ML mechanism.  Proposal 5: Study the mechanism of model selection to facilitate AI/ML model lifecycle management. |
| Oppo | Proposal 12: Give high priority to some basic LCM solutions, e.g. the selection and use of the most suitable scheme through reasonable performance monitoring, necessary signaling indication and model switching.  Proposal 13: More challenging LCM schemes, e.g. online real-time model training and updating, can be studied in subsequent studies. |
| FutureWei | Proposal 7: Study potential standards impact to support AI/ML model adaptation across various configurations. |
| CATT | Proposal 5: Study mechanisms on model quality monitoring for CSI feedback, with the following aspects considered:  – Which side takes responsibility on model quality monitoring, e.g. at UE side, at network side, or both;  – The scheme of model quality monitoring when only partial of AI/ML model (i.e. encoder/decoder) is known by one side. |
| NVIDIA | Proposal 6: For AI/ML based CSI feedback, study potential specification impact related to assistance ignaling and procedure for model configuration, model activation/deactivation, model recovery/termination, and model selection.  Proposal 7: For AI/ML based CSI feedback, study potential specification impact related to assistance ignaling and procedure for model performance monitoring and model update/tuning. |
| Spreadtrum Communications | Proposal 6: Both gNB and UE can be considered to monitor AI/ML model.  Proposal 7: The better generalization of AI/ML model should be strived, to avoid frequent AI/ML model updating. |
| Xiaomi | Proposal 4: It should be studied which side implementing performance monitoring or what is the metric of performance monitoring. |
| CAICT | Proposal 4: Some original CSI information feedback could be considered for AI/ML model monitoring. |
| Samsung | Proposal 2-3: Study and verify model update of the encoder at the UE, where the gNB’s training strategy is not disclosed while transferring/configuring the AE |
| LGE | Proposal #4: Consider fallback operation when AI/ML based CSI reporting is not valid. |
| CMCC | Proposal 7: For AI based CSI enhancement, the potential spec impact of model selection/model switching should be studied. |
| Interdigital | Proposal 10: Study means to monitor the AI/ML encoder performance at inference time, for CSI enhancements using both two-sided AI/ML model, and one-sided AI/ML at the UE.  Proposal 11: Study means to mitigate AI/ML encoder model performance degradation.  Proposal 12: Study means to perform model transfer to mitigate AI/ML model performance degradation  Proposal 13: Mechanisms to fallback to legacy CSI reporting are needed (e.g. for cases where model transfer or AI/ML model on-line training is performed) |
| Qualcomm | Proposal 12: Study signaling and procedures for X-node CSF performance monitoring, including assistance information, performance report and indication of model deactivation, retraining or switching. |
| Apple | Proposal 8: Activation/de-activation/switching can be enabled by RRC configuration.  Proposal 9: Performance monitoring can be done at the UE and the gNB based on DL throughput or PDSCH BLER. |
| NTT DOCOMO | Proposal 4: Study NW-based model monitoring and UE-based model monitoring for spatial-frequency domain CSI compression. |

## ***Proposal 3-3-4:***

***In CSI compression using two-sided model use case, study potential specification impact for performance monitoring, considering at least the following aspects:***

* ***NW-based AI model performance monitoring and UE-based AI model performance monitoring***
* ***Performance monitoring KPIs***
* ***Assisted information and performance report***
* ***Co-existence and fall-back mechanisms between AI/ML mode and legacy non-AI/ML mode.***

Please provide your view below:

|  |  |
| --- | --- |
| **Company** | **View** |
| Lenovo | Prefer to focus on prior proposals first. Performance monitoring can be discussed at a later stage |
| OPPO | support |
| CAICT | Support |
| Futurewei | We are ok with the proposal. |
| MediaTek | OK to discuss but we can wait for discussion for general framework in AI 9.2.1 |
| NVIDIA | Support |
| Xiaomi | Support |
| ZTE | We generally agree with this proposal, and we think a sub-bullet **Self-monitoring** can be added for supplementary.  The details about self-monitoring are introduced in our proposal of 9.2.1.   * An AI/ML model is able to self-monitor its performance, which is the most convenient and efficient way since no additional RS overhead or report overhead is required. For example, if the AI/ML model can detect that the distribution between training dataset and inference data has been changed a lot, it cannot make sure that the AI/ML model can get expected output. |
| Google | Support |
| Samsung | Support |
| NTT DOCOMO | Support the proposal. It is important to monitor the (near) real time model performance for reliable NW operations. As performance monitoring process could be different according to sub use case, the performance monitoring specific to CSI compression using two-sided model use case should be studied. |
| Qualcomm | Performance monitoring could be direct (e.g., inference accuracy) or indirect (e.g., system throughput). Indirect monitoring could be NW-based or UE-based. The need for direct monitoring is FFS, and if there is a need, then the mechanisms for direct monitoring are also FFS. |
| Ericsson | Support in principle but this is being discussion in 9.2.1 as well. The study here should focus on whether there are any CSI specific performance monitoring needs with specification impact. |
| LG | Support |
| vivo | We are fine with the proposal to set a starting point for the discussion of performance monitoring. |
| CMCC | Support. |
| CATT | Support |
| Spreadtrum | Support |
| Fujitsu | We support this proposal in general In addition, we suggest adding “signal to generate label for monitoring” to the second bullet.   * *Performance monitoring KPIs and the signal to generate label for monitoring.* |
| Sony | Support |

## ***Proposal 3-3-5:***

***In CSI compression using two-sided model use case, further study potential specification for model selection, model configuration, model activation/de-activation, model adaptation across various configurations.***

Please provide your view below:

|  |  |
| --- | --- |
| **Company** | **View** |
| Lenovo | Prefer to focus on prior proposals first. Performance monitoring can be discussed at a later stage |
| OPPO | support |
| CAICT | Support |
| Futurewei | We are ok with the proposal. |
| MediaTek | OK to discuss but we can wait for discussion for general framework in AI 9.2.1 |
| NVIDIA | Support |
| Xiaomi | Support |
| ZTE | Support in general. |
| Google | Support |
| Samsung | Support |
| NTT DOCOMO | Support the proposal. Adequate LCM could be different according to sub use cases. LCM specific to CSI compression using two-sided model use case should be studied. |
| Qualcomm | Support. We suggest adding “model switching”. In case multiple models are needed, the multiple models can be developed offline, stored at a server, and registered to the network with model IDs, and model switching can be used to switch among the multiple models during inference. It is not clear what “model adaptation” means and why it is needed. |
| Ericsson | Support in principle but this is being discussion in 9.2.1 as well. The study here should focus on whether there are any CSI enhancement related needs with specification impact. |
| LG | Support |
| vivo | Support |
| CMCC | Support. |
| CATT | Support |
| Spreadtrum | Support |
| Fujitsu | Support. |
| Sony | Support |

## Model transfer

Following table summarize company’s proposals related to model transfer.

|  |  |
| --- | --- |
| Vivo | Proposal 11: Study the specification impacts of following model transfer cases:  • Option 0: Network sends updated parameters and does not change the AI/ML model structure. (Already supported by nowadays typical chipset implementations)  • Option 1: Network sends AI/ML model parameter and structure information. (Dependent on how much the model structure is changed, recompilation may be needed.)  Proposal 12: Further study model transfer format in both short and long term. We suggest to consider the following options:  • Negotiation between networks nodes  • Definition of a common format recognizable by multiple parties  • Definition of a new model description format by 3GPP |
| Futurewei | Proposal 5: For solution options required model transfer, study potential standards impact associated with:  o Exchanging model information, including protocol/ignaling mechanism that enables the model transfer  o Exchanging additional functional modules (if not integrated with the model) and/or other supporting information between gNB and UE to help the receiving node to perform the encoding/decoding function and/or interpret the model  Proposal 6: Study potential standards impact related to deploying/supporting different encoders based on UE capability while keeping the main decoder architecture unchanged at gNB. |
| CMCC | Proposal 4: For AI based CSI enhancement, the potential spec impact on AI model transfer need to be studied. |
| MediaTek | Proposal 7: Discuss potential spec impact on model exchange focusing on the followings  • Content of the model exchange including model format, pre/post-processing choice, model parameters, hyper-parameters, etc.  • Signalling format for the model exchange  • Related UE capability |
|  |  |

Model exchange can be discussed at least initially at 9.2.1. CSI specific model exchange will be discussed later after high level discussion in 9.2.1.

## Framework, UE capability, and other topics

Following table summarize company’s proposals related to framework.

|  |  |
| --- | --- |
| Google | Proposal 1: The study of AI/ML based CSI compression should be based on the CSI framework in Rel-17.  Proposal 3: The study of the report of compressed CSI should be based on Rel-15 CSI report mechanism, where the CSI is reported in a single part in short PUCCH, and the CSI can be reported in two parts in long PUCCH and PUSCH.  Proposal 4: Study the priority rule for AI/ML based CSI report and non-AI/ML based CSI report with regard to CSI collision handling and CSI omission.  Proposal 5: The AI/ML based CSI compression should consider the following types of UE:  • Type 1 UE (low performance UE): CSI compression is based on general processing unit (GPU)  • Type 2 UE (high performance UE): CSI compression is based on neural processing unit (NPU) |
| Oppo | Proposal 6: Discussions on AI/ML model deployment and AI/ML model training should be decoupled.  Proposal 7: In Rel-18, fully analyze the difficulty and requirement of AI/ML model deployment, and distinguish the impact of different conditions and assumptions, including:  - Real-time deployment  - Non real-time deployment  - Whole new model deployment  - Partial new model deployment (e.g. the deployment of only updating model weights)  - Deployment of complex models  - Deployment of simple models  Proposal 8: Scenarios for non real-time, partial model deployment and simple model deployment can be considered as the basic deployment assumption for subsequent research in Rel-18.  - FFS Other scenarios  - Scenarios with high deployment complexity can wait until it could be handled by companies with reasonable solutions  Proposal 9: Both protocol visible interfaces and protocol invisible interfaces can be used in subsequent AI/ML applications and need to be studied. |
| CATT | Proposal 6: Study the scalable and flexible frameworks for AI/ML based approaches for CSI feedback.  Proposal 7: On evaluation of scalability of AI/ML model for CSI feedback, the following configurations can be considered as the starting point:  - Different number of antenna ports, e.g. 32 ports, 16 ports.  - Different number of reporting subbands.  - Different bandwidths, e.g. 10MHz vs. 20MHz.  - Different numerologies, e.g. 15kHz vs. 30kHz.  - Different CSI feedback payloads. |
| Spreadtrum Communications | Proposal 1: Legacy CSI framework can be reused for the sub use case - Spatial-frequency domain CSI compression. Additional enhancement can be considered.  Proposal 5: How to define/reflect the complexity of the AI/ML operation in the specification should be considered. |
| Samsung | Proposal 4-1: Study UE processing time impact on online training for update, transfer, and download. |
| LGE | Proposal #3: Consider enhancement of UE CSI processing procedure including CPU and CSI reference resource for AI/ML based CSI reporting. |
| ETRI | Proposal 2: Study the feasibility of the PCA based two-sided AI model for CSI compression sub use case. |

UE capability and other topics will be discussed later. High level agreement on framework can be benifical for future discussion.

## ***Proposal 3-6:***

***The study of AI/ML based CSI compression should be based on the legacy CSI feedback framework.***

Please provide your view below:

|  |  |
| --- | --- |
| **Company** | **View** |
| Lenovo | This proposal is not clear to us. Legacy CSI feedback, e.g., Rel-16 Type-II CB can be used as a baseline for throughput/overhead/complexity comparison, but the AI-based CSI compression scheme should not be based on legacy codebook design |
| MediaTek | Similar views with Lenovo. What is the intention for this proposal? |
| NVIDIA | It’s not clear what this proposal aims to achieve. |
| Google | Support. We do not need to introduce a new framework for AI/ML based CSI feedback. Current study on spec impact related aspects should be based on legacy CSI framework. |
| Samsung | Could you clarify “legacy CSI feedback framework”? Does that refer to Rel-16 enhanced Type II, Rel-17 FeType II PS, etc.? |
| NTT DOCOMO | Fine with the proposal. The legacy CSI feedback/framework can be reused, if possible. |
| Qualcomm | Support the principle, but the meaning of the proposal should be clarified. It is not clear what aspects of the feedback framework are being referred to here. |
| Ericsson | Unclear what the proposal means…??. |
| LG | Support |
| vivo | Agree with Lenovo’s comments. Clarifications on what “legacy CSI feedback framework” refers to is required before further discussion. |
| CMCC | Not very clear on the “legacy CSI feedback framework”. |
| CATT | It is not clear what ‘***legacy CSI feedback framework***’ mean here. |
| Spreadtrum | Support. |
| Fujitsu | No need to add such restriction at this stage. |

# Potential specification impact for other use cases

The following table summarizes proposals on potential specification impact on other sub-use cases. The summary can be used as reference for discussion of section 2.

|  |  |
| --- | --- |
| vivo | Proposal 15: Study the specification impact of both gNB- and UE-based CSI prediction.  Proposal 16: For UE-based CSI prediction, study on specification impact at least includes the following aspects  • Capability report of CSI prediction  • gNB’s activation, deactivation, configuration and adjustment of AI based CSI prediction, and UE’s request on such actions  • gNB and UE’s alignment on prediction related time domain configuration information  • Supported CSI-RS configurations (e.g., CSI-RS time domain type(s))  • Correct CSI reference resource definition  Proposal 17: To support gNB-based prediction with high accuracy, the CSI feedback enhancement should be carefully designed to reserve Doppler information or time varying information as much as possible.  Proposal 18: Study on LCM aspects of CSI prediction at least includes the following  • For performance monitoring, functionality of using dedicated CSI-RS and reporting process to derive label with lower noise and interference  • LCM of chained AI model (e.g., for AI-based prediction and compression)  • Finetuning process of AI-based CSI prediction |
| Fujitsu | Proposal-4: For the one-sided AI/ML-based method, the study of STD impacts may start from:   Studying the time-domain CSI-RS configurations to enable accurate CSI prediction.   Studying CSI-RS overhead reduction for CSI prediction.   Studying assistance information needed for CSI prediction. |
| NEC | Proposal 6: Support the location/CQI report timing set mapping table based on AI/ML.  Proposal 7: Support the location/CQI periodicity mapping table based on AI/ML. |
| Apple | Proposal 10: For CSI prediction use case, potential specification impact including UE capability signaling, UE request and NW activation/de-activation signaling. |
| Interdigital | Proposal 9: Specification impacts of reduced CSI-RS overhead include new RS configurations, new RS triggers, and UE feedback. |

# Reference

[R1-2205891](file:///C:\Users\younsun\Documents\3GPP%20documents\RAN1%20tdocs\TSGR1_110\Docs\R1-2205891.zip) Discussion on AI/ML for CSI feedback enhancement Huawei, HiSilicon

[R1-2205967](file:///C:\Users\younsun\Documents\3GPP%20documents\RAN1%20tdocs\TSGR1_110\Docs\R1-2205967.zip) Discussions on Sub-Use Cases in AI/ML for CSI Feedback Enhancement TCL Communication

[R1-2206033](file:///C:\Users\younsun\Documents\3GPP%20documents\RAN1%20tdocs\TSGR1_110\Docs\R1-2206033.zip) Other aspects on AI/ML for CSI feedback enhancement vivo

[R1-2206069](file:///C:\Users\younsun\Documents\3GPP%20documents\RAN1%20tdocs\TSGR1_110\Docs\R1-2206069.zip) Discussion on other aspects for AI CSI feedback enhancement ZTE

[R1-2206114](file:///C:\Users\younsun\Documents\3GPP%20documents\RAN1%20tdocs\TSGR1_110\Docs\R1-2206114.zip) Considerations on CSI measurement enhancements via AI/ML Sony

[R1-2206165](file:///C:\Users\younsun\Documents\3GPP%20documents\RAN1%20tdocs\TSGR1_110\Docs\R1-2206165.zip) Discussion on other aspects of AI/ML for CSI feedback enhancement Fujitsu

[R1-2206185](file:///C:\Users\younsun\Documents\3GPP%20documents\RAN1%20tdocs\TSGR1_110\Docs\R1-2206185.zip) Discussion on AI/ML for CSI feedback enhancement Panasonic

[R1-2206196](file:///C:\Users\younsun\Documents\3GPP%20documents\RAN1%20tdocs\TSGR1_110\Docs\R1-2206196.zip) On Enhancement of AI/ML based CSI Google

[R1-2206241](file:///C:\Users\younsun\Documents\3GPP%20documents\RAN1%20tdocs\TSGR1_110\Docs\R1-2206241.zip) Discussion on AI/ML for CSI feedback enhancement NEC

[R1-2206316](file:///C:\Users\younsun\Documents\3GPP%20documents\RAN1%20tdocs\TSGR1_110\Docs\R1-2206316.zip) On sub use cases and other aspects of AI/ML for CSI feedback enhancement OPPO

[R1-2206337](file:///C:\Users\younsun\Documents\3GPP%20documents\RAN1%20tdocs\TSGR1_110\Docs\R1-2206337.zip) Continued discussion on other aspects of AI/ML for CSI feedback enhancement FUTUREWEI

[R1-2206392](file:///C:\Users\younsun\Documents\3GPP%20documents\RAN1%20tdocs\TSGR1_110\Docs\R1-2206392.zip) Other aspects on AI/ML for CSI feedback CATT

[R1-2206511](file:///C:\Users\younsun\Documents\3GPP%20documents\RAN1%20tdocs\TSGR1_110\Docs\R1-2206511.zip) Further aspects of AI/ML for CSI feedback Lenovo

[R1-2206521](file:///C:\Users\younsun\Documents\3GPP%20documents\RAN1%20tdocs\TSGR1_110\Docs\R1-2206521.zip) AI and ML for CSI feedback enhancement NVIDIA

[R1-2206579](file:///C:\Users\younsun\Documents\3GPP%20documents\RAN1%20tdocs\TSGR1_110\Docs\R1-2206579.zip) Use-cases and specification for CSI feedback Intel Corporation

[R1-2206605](file:///C:\Users\younsun\Documents\3GPP%20documents\RAN1%20tdocs\TSGR1_110\Docs\R1-2206605.zip) Discussion on other aspects on AIML for CSI feedback Spreadtrum Communications

[R1-2206636](file:///C:\Users\younsun\Documents\3GPP%20documents\RAN1%20tdocs\TSGR1_110\Docs\R1-2206636.zip) Discussion on potential specification impact for CSI feedback based on AI/ML Xiaomi

[R1-2206676](file:///C:\Users\younsun\Documents\3GPP%20documents\RAN1%20tdocs\TSGR1_110\Docs\R1-2206676.zip) Discussions on AI-ML for CSI feedback CAICT

[R1-2206687](file:///C:\Users\younsun\Documents\3GPP%20documents\RAN1%20tdocs\TSGR1_110\Docs\R1-2206687.zip) Discussion on AI/ML for CSI feedback enhancement China Telecom

[R1-2206821](file:///C:\Users\younsun\Documents\3GPP%20documents\RAN1%20tdocs\TSGR1_110\Docs\R1-2206821.zip) Representative sub use cases for CSI feedback enhancement Samsung

[R1-2206875](file:///C:\Users\younsun\Documents\3GPP%20documents\RAN1%20tdocs\TSGR1_110\Docs\R1-2206875.zip) Other aspects on AI/ML for CSI feedback enhancement LG Electronics

[R1-2206884](file:///C:\Users\younsun\Documents\3GPP%20documents\RAN1%20tdocs\TSGR1_110\Docs\R1-2206884.zip) Discussion on AI-CSI Ericsson

[R1-2206903](file:///C:\Users\younsun\Documents\3GPP%20documents\RAN1%20tdocs\TSGR1_110\Docs\R1-2206903.zip) Discussion on other aspects on AI/ML for CSI feedback enhancement CMCC

[R1-2206954](file:///C:\Users\younsun\Documents\3GPP%20documents\RAN1%20tdocs\TSGR1_110\Docs\R1-2206954.zip) Discussion on other aspects on AI/ML for CSI feedback enhancement ETRI

[R1-2206969](file:///C:\Users\younsun\Documents\3GPP%20documents\RAN1%20tdocs\TSGR1_110\Docs\R1-2206969.zip) Other aspects on ML for CSI feedback enhancement Nokia, Nokia Shanghai Bell

[R1-2206989](file:///C:\Users\younsun\Documents\3GPP%20documents\RAN1%20tdocs\TSGR1_110\Docs\R1-2206989.zip) Other aspects on AI/ML for CSI feedback enhancement MediaTek Inc.

[R1-2207153](file:///C:\Users\younsun\Documents\3GPP%20documents\RAN1%20tdocs\TSGR1_110\Docs\R1-2207153.zip) Discussion on AI/ML for CSI feedback enhancement InterDigital, Inc.

[R1-2207225](file:///C:\Users\younsun\Documents\3GPP%20documents\RAN1%20tdocs\TSGR1_110\Docs\R1-2207225.zip) Other aspects on AI/ML for CSI feedback enhancement Qualcomm Incorporated

[R1-2207329](file:///C:\Users\younsun\Documents\3GPP%20documents\RAN1%20tdocs\TSGR1_110\Docs\R1-2207329.zip) Other aspects on AI/ML for CSI Apple

[R1-2207370](file:///C:\Users\younsun\Documents\3GPP%20documents\RAN1%20tdocs\TSGR1_110\Docs\R1-2207370.zip) Sub-use cases for AI/ML feedback enhancements AT&T

[R1-2207402](file:///C:\Users\younsun\Documents\3GPP%20documents\RAN1%20tdocs\TSGR1_110\Docs\R1-2207402.zip) Discussion on other aspects on AI/ML for CSI feedback enhancement NTT DOCOMO, INC.