**3GPP TSG RAN WG1 #122 R1- 250xxxx**

**Bengaluru, India, Aug 25th – 29th, 2025**

Agenda Item: 10.1.2

Source: Moderator (Apple)

Title: FL summary # 1 for inter-vendor training collaboration

Document for: Discussion/Decision

# Introduction

The objective on inter-vendor training collaboration approved in RP-251870 is as follows.

**Inter-vendor training collaboration** for two-sided AI/ML models

* Fully defined/specified reference model (“Direction C”) with RAN1 scalability study outcome taken into account [RAN4/RAN1] – check-point in RAN#110 upon RAN4 feedback
	+ Specification of standardized encoder model structure plus parameter exchange (“Direction A, sub-option 3a-1” without target CSI sharing) leveraging defined/reference model of “Direction C” and taking RAN1 scalability study outcome into account [RAN4/RAN1/RAN2/RAN3] – check-point in RAN#110 upon SA WG feedback
* Specification of standardized dataset format/content plus dataset exchange (“Direction A, sub-option 4-1”) [RAN1/RAN2/RAN3/RAN4] – check-point in RAN#110 upon SA WG feedback

This document summarizes the contributions in RAN1 #122 on agenda 10.1.2.

## Contact information

## Please provide your contact information.

|  |  |  |
| --- | --- | --- |
| Company | Name | Email |
| Moderator (Apple) | Huaning Niu | huaning\_niu@apple.com |
| OPPO | Wendong Liu | liuwendong1@oppo.com |
| Lenovo | Vahid Pourahmadi | vpourahmadi@lenovo.com |
| NTT DOCOMO | Xin Wang | wangx@docomolabs-beijing.com.cn |
| Huawei, HiSilicon | Keyvan Zarifi, Yuan Li | Keyvan.zarifi@huawei.com, liyuan3@huawei.com |
| Ericsson | Jingya Li, Xinlin Zhang | Jingya.li@ericsson.comXinlin.zhang@ericsson.com |
| MediaTek | Pedram Kheirkhah SangdehReubengeorge Stephen | Pedram.kheirkhah@mediatek.comReubengeorge.stephen@mediatek.com |
| ZTE | Hanchao LiuWenfeng Liu | liu.hanchao@zte.com.cnliu.wenfeng@zte.com.cn |

# 2 Discussion plan

Based on R1-2506206, the rapporteur's work plan, and cross agenda guidance on overlapping topics, the following high-level plan is summarized as reference.

***Discussion on inter-vendor training collaboration options***

Inter-vendor training collaboration for two-sided model includes 3 options, “Direction A, sub-option 4-1”, “Direction A, sub-option 3a-1”, “Direction C”. Among these, two options are led by RAN4, while one is led by RAN1.

According to rapporteur's work plan, RAN1 should focus on Direction A, sub-option 4-1. In parallel, RAN4 is expected to advance work on Direction C, considering the scalability study results from RAN1. RAN4 will also explore Direction A, sub-option 3a-1 without targeted CSI sharing, leveraging work done under Direction C. Therefore, collaboration between RAN1 and RAN4 is anticipated.

Based on proposal 3 of R1-2506206, discussion in agenda 10.1.2 will start with “Direction A, sub-option 4-1” in this meeting.

**“Proposal 3**: RAN1 focus should be on specification of standardized dataset format/content plus dataset exchange (“Direction A, sub-option 4-1”). RAN1 should assist RAN4 with their work when requested.”

***Overlapping discussion with 10.1.1***

Several topics may overlap with agenda items 10.1.1.1 and 10.1.1.2. To improve discussion efficiency, the following high-level plan will be followed:

* **Target CSI Type/Format**:
	+ *Target CSI type* (channel or precoder, antenna-port-subband representation or angular-delay representation)will be discussed under **10.1.1.1**.
	+ *Target CSI Format* (e.g., scalar quantization or e-type 2 codebooks) is a method to quantize the target CSI and deliver it from UE to NW in data collection, performance monitoring or exchanged from NW to UE for inter-vendor collaboration. It overlaps with **10.1.1.2**. Initial discussion focus in **10.1.2** is whether the same format or a different format should be used compared to 10.1.1.2.
* **Pairing ID / Dataset ID**:
Pairing ID is relevant for training collaboration, data collection, and inference pairing. **10.1.2** plans to discuss paring ID as part of the dataset content topic.
* **Quantization Codebook**:
Quantization-aware training is the baseline assumption for training collaboration. It is included as part of the dataset exchange for training in **10.1.2**.

# 3 Summary and proposals

## 3.1 Target CSI and CSI feedback

It is proposed that both the target CSI and CSI feedback be included as part of the dataset. For CSI feedback, there are differing views on its role in the quantization process:

* If CSI feedback is defined as the floating-point values at the input of quantization, better performance has been observed.
* Alternatively, defining CSI feedback as the binary sequence after quantization reduces overhead.

### In addition, association between target CSI and CSI feedback has been discussed. It can be 1:1 mapping. In this case, to support scalability across CSI payload size, sub bands and antenna ports, the number of data samples can be very large. It was also proposed as 1:M mapping, to support different CSI feedback size and subbands.

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

***For Option 4-1 under Direction A in AI/ML based CSI compression, support at least target CSI and CSI feedback in the standardized dataset.***

* ***FFS: Target CSI type and format***
* ***FFS: CSI feedback type and format***
* ***FFS: Association between Target CSI and CSI feedback, including scalability related information for for different number of Tx port, number of sub bands, and CSI payload size.***

Please provide your view below:

|  |  |
| --- | --- |
| **Company** | **View** |
| OPPO | Support |
| Lenovo | Support |
| NTT DOCOMO | Support |
| Huawei, HiSilicon | Support |
| Xiaomi | Support |
| Ericsson | Support |
| MediaTek | Support |
| ZTE | Support in general Per our understanding, the dataset should be the exchanged dataset rather than a standardized one. So, we suggest the following wording:***For Option 4-1 under Direction A in AI/ML based CSI compression, support at least target CSI and CSI feedback in the ~~standardized~~ exchanged dataset.*** For the last bullet, we seek clarification on term 'scalability related information', whether it specifically denotes the mapping relationship between target CSI under various configurations and CSI feedback under various configurations. |

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

***For Option 4-1 under Direction A in AI/ML based CSI compression, for CSI feedback type and format, down select one of the following options:***

* ***Option 1: CSI feedback is defined as the floating-point values at the input of quantization.***
* ***Option 2: CSI feedback is defined as the binary sequence at the output of quantization.***

Please provide your view below:

|  |  |
| --- | --- |
| **Company** | **View** |
| OPPO | Fine and prefer option 2 after quantization. |
| Lenovo | As the UE needs to train the encoder model, it is beneficial if it can have access to the floating-point values of the output of the “nominal(reference) Encoder” at the gNB side. Such information may help better alignment between the “nominal (reference) encoder” model and the encode that the UE-side will develop. So, we prefer Option-1 and we note that the quantization scheme will be communicated separately between the NW-side and the UE-side so, the UE-side can also generate the quantized version itself if needed.  |
| NTT DOCOMO | Option 2 is preferred based on the previous study on the quantization-aware training. |
| Qualcomm | Both of them are needed, or option 1 + codebook. We are not sure how encoder can be trained without option 1 – I wonder option2 proponents assume UE would directly output the quantized version? |
| Huawei, HiSilicon | Support and prefer Option-1. Based on Rel-18 studies, both options are functional while Option-1 showed slightly better SGCS performance. We prefer Option-1 both based on the SGCS performance and the fact that Option-1 provides more information to the UE side. Note that the quantization details should be provided to the UE side in either of the Options anyway and, if the floating-point values are provided to the UE side (option 1), UE can directly rebuild the quantized output as in Option-2. Since the inter-vendor exchange uses non-OTA mechanism, the feedback overhead associated with Option-1 is not a concern. |
| Xiaomi | In our view, for option 1, quantization approach should be known by the UE side. One possible approach is to fix it in the spec and another approach is to include it in the data delivery. If only one quantization manners is supported, then we think both options are equivalent. If multiple quantization manners are supported, it seems option 2 is better as Lenovo commented. Considering this aspect, we think it depends on the conclusion of the quantization manners. We could come back this proposal when there is progress for the quantization manner. |
| Ericsson | Based on the agreement made in Rel-19, quantization-related parameters will be exchanged from NW side to UE side along with each exchanged dataset for option 4-1. We don’t see big differences between option 1 and option 2 from the e2e performance perspective.For option 2, the CSI feedback included for dataset exchange shares the same format as the one defined for inference, this can reduce the additional spec impact, and reduce the dataset size, especially considering that the number of data samples for dataset exchange can be large. Hence, option 2 is preferred. |
| MediaTek | We prefer option 1 as it provide slightly better performance due to potential error propagation from quantization into the training of the encoder. Also, our intention has never been to merge quantization in to the model and quantization codebook or parameter will be shared by NW; therefore, we do not see any motivation to pursue option 2.  |
| ZTE | Prefer option 2. Compared with Option 1, Option 2 can reduce the dataset size and transmission overhead.​ |
|  |  |

For target CSI format, there are proposals discussing reuse or define target CSI format separately comparing to NW side data collection for training. Several different aspects are discussed:

* NW side data collection needs to consider UE complexity/overhead aspects, while dataset delivery via non-OTA has not this concern considerations.
* UE generates e-type 2 PMI is part of UE legacy MIMO operation and has RAN4 test to ensure PMI is generated accurately. However, training entity to generate the e-type 2 PMI is an additional requirement, and how to ensure accuracy is not clear.
* For scalability model training, e-type 2 PMI with different sub band size will have different PMI coefficients. Related to association between target CSI versus CSI feedback, whether it is 1:1 or 1:M, target CSI format can have different impact.

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

***For Option 4-1 under Direction A in AI/ML based CSI compression, for target CSI format, further study the following two approaches:***

* ***Option 1: Target CSI format reuses the same format as NW-side data collection for training.***
* ***Option 2: Target CSI format for inter-vendor training collaboration is designed separately.***

Please provide your view below, particularly the reasoning for support option 1 or option 2:

|  |  |
| --- | --- |
| **Company** | **View** |
| OPPO | Fine with this proposal. Regarding option 1 and option 2, we have following considerations:For the format of NW-side data collection, if only one kind of format is specified and supported, then every dataset reported by different UEs/UE-vendors would has the same format, thus Option 1 can be considered for simplicity.Otherwise, if multiple kinds of formats are specified and supported in NW-side data collection, Option 1 seems not workable since UE does not know which kind of format is used in target CSI format. Hence, we think Option 2 may be better. Whether to use a unified target CSI format in dataset exchange or multiple formats can be supported can be further discussed.  |
| Lenovo | We believe the “Target CSI” samples that will be shared 4or Option 4-1 under Direction A are the same samples that has been collected by the NW during the data collection phase and used for training of the Decoder model. So, the NW-side does not need to generate “Target CSI” samples again and so there is no need for specification of separate “Target CSI” format, and so we prefer Option-1. |
| Huawei, HiSilicon | Support the proposal and prefer Option-2. We think the labels for UE side model training should have a high accuracy as the accuracy of the whole trained model is upper-bounded by the accuracy of its labels. If we go with Option-1, the lower resolution Target CSI that is sent over-the-air to the NW side during NW data collection could be the bottleneck for the UE side training accuracy. Also, in principle, we don’t see a technical necessity to use the same Target CSI format for NW-side data training and inter-vendor exchange. Since inter-vendor exchange is non-OTA, feedback overhead is not a concern and a high-resolution format (eg, FP32) may be considered.  |
| Xiaomi | At current stage, we slightly tend to de-couple the data format between the network data collection and the inter-vendor training collaboration. In our understanding, it is maybe OK to collect raw data during the data collection phase, it allows the network to have more freedom do the post-processing to leverage different target CSI format in future.  |
| Ericsson | Support Option 1. We think both NW-side data collection and dataset exchange from NW-side to UE-side shall be done via higher layer signaling. And the same high-resolution target CSI format shall be supported for both purposes.Option 2 may result in increased inter-vendor training collaboration complexity (e.g., when different vendors support different formats), which is the problem we try to resolve on this agenda item. |
| MediaTek | Support option1 for ease of inter-vendor collaboration and avoid any conversion between different formats for training.  |
| ZTE | We prefer Option 1 because it reduces the workload on both the network side and the standardization process. |
| vivo | For option 1, we have a question: Does reusing the same format mean that all of the CSI part 1 information must be exchanged with the PMI? Specifically, does the CQI or rank information need to be transmitted to the UE? From our perspective, data grouped by layer is more logical than mixing data with different ranks, since a layer common model is agreed upon in the Rel-19 study. Therefore, could we consider selecting only CSI part 2 and the total number of non-zero coefficients for the target CSI, and organizing all target CSI by layer |

## 3.2 Assisted information for scalable model training

Performance metrics are proposed as part of assisted information to facilitate the model training. Two metrics, SGCS and NMSE, were studied. SGCS is the metric when UE-side first trains a nominal decoder, then develop the actual encoder against the nominal decoder. NMSE is used when UE-side develops the actual encoder directly. It should be noted that the conventional NMSE definition is suitable when the target CSI is defined as floating-point value before quantization. However, for binary target CSI, a new NMSE definition was proposed. In subsequent proposals, the new version has been renamed to avoid confusion with the traditional NMSE.

In Rel-19, scalable model and scalable model structure has been studied to support various subbands, ports and payload configurations. Several alternatives are identified in R19 study. Assisted information for training related to scalable architecture and dimension mapping was also proposed.

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

***For Option 4-1 under Direction A in AI/ML based CSI compression, support the performance target in the standardized dataset, with a potential down-selection between the following performance target options:***

* ***Average SGCS.***
	+ ***FFS: SGCS values at X-percentiles***
* ***NMSE: when CSI feedback is defined as the floating-point values at the input of quantization***
* ***BER (bit error rate): when CSI feedback is defined as the binary bit sequence at the output of quantization***
* ***FFS: Multiple performance targets for different layer, different configurations such as antenna ports, subband configuration and payload configuration***

Please provide your view below, particularly the reasoning for support option 1 or option 2:

|  |  |
| --- | --- |
| **Company** | **View** |
| OPPO | We are support of using average SGCS per layer. In our view, the differential SGCS over the SGCS of legacy codebook can also be indicated in performance target. Since the performance gain over legacy codebook is more important.  |
| NTT DOCOMO | We have a general comment that the discussion about the performance target should be applied to both Direction A options (Option 3a-1 and Option 4-1). It is a common issue between the two options.  |
| Qualcomm | Suggest to revise the first bullet as “SGCS”. Details are to be further discussed. Just like other two terms “NMSE” and “BER”. we don’t need to say “average NMSE” or “average BER”. |
| Huawei, HiSilicon | Regarding SGCS and NMSE, we think SGCS is more appropriate if UE first train the nominal decoder before training its encoder while NMSE is more appropriate if the UE directly train its encoder. Since how UE trains its model(s) may not be specified and left to the UE implementation, it may be required to include both SGCS-based and NMSE/BER-based performance target values in the dataset exchange.Regarding BER,we need more justification about the motivation and necessity of using BER as a performance target. We thinkeEven if the CSI feedback is after quantization, since UE side is provided by the quantization details, the CSI feedback can be dequantized at the UE side and NMSE is used. Further, we don’t think “X-percentile values for SGCS” should be an FFS under average SGCS and can be an independent bullet. Average SGCS may not be a good indicator of how well the model is trained as the average SGCS value may be acceptable but the SGCS CDF have a long tail. Finally, regarding the last FFS, RAN1 has not agreed on target CSI type (precoder vs Channel matrix), discussing performance target per layer is pre-mature. We suggest the following change:* ***FFS: Multiple performance targets for different layer if the target CSI type is precoding matrix, different configurations such as antenna ports, subband configuration and payload configuration***
 |
| Xiaomi | * For the SGCS, we support to consider it as performance metric
* For the NMSE and BER, it is related to the discussion of proposal 1-2 (the CSI feedback type). If the CSI feedback type is float-based, then NMSE is more feasible. If the CSI feedback is binary sequence, then BER is better. Down-selection between these two options can be considered when the CSI feedback type is determined
 |
| Ericsson | The agreements made in Rel-19 regarding performance target are copied below. For Rel-20 WI, we think that* the discussion should focus on the identified performance target type SGCS and/or NMSE for Direction A option 4-1, without including new type like BER. The third bullet about BER in the proposal shall be removed. We think for option 4-1, only SGCS shall be supported as the performance target type.
* The format (average or distribution based) can be discussed after an agreement is made on which performance target type (SGCS and/or NMSE) to select.
* The FFS regarding input data for evaluating the performance is missing.

Agreements copied below:AgreementFor inter-vendor-collaboration Options 3a-1 and 4-1 in Direction A, performance target is confirmed as additional information along with the exchanged dataset or the model parameters.* FFS: type of performance metric
* FFS: input data for evaluating the performance

AgreementFor inter-vendor-collaboration Options 3a-1 and 4-1 in Direction A, confirm SGCS and NMSE as the type of performance metric that may be used for the performance target shared as additional information along with the exchanged dataset or the model parameters.* FFS: when to use SGCS, NMSE, and which one to use or both, and relationship with the inter-vender collaboration sub-options.
* FFS: details of the format of the performance target
	+ Option 1: Average performance target, e.g. average SGCS and/or average NMSE
	+ Option 2: distribution of the performance target, e.g., SGCS / NMSE for 5, 10, 20, 30 percentiles, etc.
* FFS: whether multiple performance targets should be exchanged for different configurations, such as antenna ports configuration, subband configuration and payload configuration, etc., along with each exchanged dataset or model parameters
 |
| ZTE | Similar to the proposal 1-1, the dataset should be the exchanged dataset rather than a standardized one.Average SGCS and SGCS values at X-percentiles can be discussed with equal priority in current stage.For the third bullet, BER is unsuitable as a performance target because even a single bit error in the bit sequence can cause the corresponding value of CSI feedback to deviate significantly from the correct one, for example, the index of selected vector may vary a lot even 1 bit-error occurs.For the last bullet, it is crucial to discuss whether each of configuration combinations of payload sizes, Tx ports, and subbands requires a separate performance target. For instance, if a dataset is associated with 12 configuration combinations of {2 Tx ports, 2 subbands, 3 payload sizes}, there will be 12 distinct performance targets for the dataset. However, the number of performance target explodes with the increasing number of total configuration combinations, which may be too much to exchange along with dataset.So, we suggest rewording the proposal as below:***For Option 4-1 under Direction A in AI/ML based CSI compression, support the performance target in the ~~standardized~~ exchanged dataset, with a potential down-selection between at least the following performance target options:*** * ***~~Average~~ SGCS.***
	+ ***~~FFS:~~ SGCS values at X-percentiles***
	+ ***Average SGCS***
* ***NMSE: when CSI feedback is defined as the floating-point values at the input of quantization***
* ***~~BER (bit error rate): when CSI feedback is defined as the binary bit sequence at the output of quantization~~***
* ***FFS: Multiple performance targets for different layer, different configurations such as antenna ports, subband configuration and payload configuration***
 |
| vivo | Sharing performance target is a common issue for option 4-1 and option 3a-1. Could we have an open discussion on this issue, considering the two options(Option 3a-1 and option 4), otherwise, where could we discuss the performance target of 3a-1 ?

|  |
| --- |
| AgreementFor inter-vendor-collaboration Options 3a-1 and 4-1 in Direction A, performance target is confirmed as additional information along with the exchanged dataset or the model parameters. |

Therefore, we propose modifying the main bullet as following:For the performance target sharing ~~Option 4-1 under~~ of Direction A in AI/ML based CSI compression,  ~~support the performance target in the standardized dataset,~~ with a potential down-selection between the following performance target options:  |

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

***For Option 4-1 under Direction A in AI/ML based CSI compression, further study the following assisted information to align the model design aspects:***

* ***Model backbone type for reference encoder, as well as hyper parameters if needed***
* ***Tokenization dimension and feature dimension mapping in the reference encoder***
* ***Scalability options used in reference encoder.***

Please provide your view below, particularly the reasoning for support option 1 or option 2:

|  |  |
| --- | --- |
| **Company** | **View** |
| OPPO | Fine to further discuss this proposal. |
| Lenovo | Fine with the proposal, but the second and third bullet are not clear to us. We suggest removing them for now. |
| NTT DOCOMO | General fine with the proposal. This discussion is also highly related to the Option 3a-1. We propose to start the works for Option 3a-1 now and discuss these issues there. |
| Qualcomm | Support.Whether these information can be exchanged or the dataset is by default to be generated assuming the RAN4 model structure. (similar to the flavor of target CSI sharing + standardized model design aspects). We are open to discuss them. |
| Huawei, HiSilicon | We support studying the last bullet only. Model backbone is a proprietary information of the vendor and should not be shared. Also, simulations during SI showed that using two different backbones at the UE side and NW side does not meaningfully impact the performance. Our understanding from companies’ evaluation results is that companies consider sub-bands as Token and number of Tx antennas as the feature. Further studying Token/feature dimensions should have a clear justification. |
| Xiaomi | Fine for further study  |
| Ericsson | We don’t see the need to exchange model design related information (e.g., model backbone type, tokenization dimension, feature dimension mapping, etc.) as additional information for option 4-1. As indicated by the study results captured in TR 38.843, the impact of different backbones at UE-part model and NW-part model is very small, and the impact is much less than the impact caused by data distribution mismatch. In addition, sharing model related information from NW-side to UE-side risks NW-side disclose proprietary implementation related information, which is a controversial topic already discussed in Rel-19, and many NW vendors do not support this. Regarding the third bullet, further clarification is needed, are the scalability options related to payload configurations? Assuming that payload size configuration and quantization parameters will be shared from NW-side to UE-side along with the exchanged dataset, what additional information is missing? |
| MediaTek | Support the exchange of all three information |
| ZTE | Regarding these three bullets, there is a potential risk of disclosing NW-side proprietary information.So, we suggest rewording the proposal as below:***For Option 4-1 under Direction A in AI/ML based CSI compression, further study the following assisted information to align the model design aspects:*** * ***Model backbone type for reference encoder, as well as hyper parameters if needed***
* ***Tokenization dimension and feature dimension mapping in the reference encoder***
* ***Scalability options used in reference encoder.***
* ***Note: NW-side proprietary information should not be disclosed.​***
 |

## 3.3 Paring ID

Paring ID has been proposed to be sent together with the dataset. Different design consideration has been proposed for paring ID, including:

* Pairing ID is generated by the training entity
* Pairing ID is unique per PLMN
* Pairing ID design should support model update by augmenting existing datasets, such as a time-stamp or version number is included as part of paring ID.
* Paring ID is associated with different scalability configurations
* Pairing ID is associated with different quantization codebook, if needed.

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

***For Option 4-1 under Direction A in AI/ML based CSI compression, support paring ID in the standardized dataset.***

* ***FFS: paring ID is PLMN unique***
* ***FFS: the association of pairing ID with different model scalability configurations***
* ***FFS: the association of paring ID with different quantization codebooks***
* ***FFS: the association of pairing ID between different datasets to enable model update***

Please provide your view below, particularly the reasoning for support option 1 or option 2:

|  |  |
| --- | --- |
| **Company** | **View** |
| OPPO | Fine with this proposal. |
| Lenovo | Fine with this proposal with the following changes.***For Option 4-1 under Direction A in AI/ML based CSI compression, support exchange of paring ID along with dataset exchange ~~in the standardized dataset.~~*** This clarifies that the dataset itself is set of samples of ***at least target CSI and CSI feedback***which will be accompanied with some other information such as pairing ID. |
| NTT DOCOMO | Fine with the proposal, and the pairing ID should be PLMN unique. |
| Huawei, HiSilicon | Some further clarification/explanation for the third FFS may be necessary to give direction to the subsequent discussions. Does it mean multiple quantization codebooks are included in the same dataset? If yes, each of these quantization codebooks are applied to which element in the dataset (eg, different codebooks for different CSI feedback sizes for the same target CSI,…) and why? |
| Xiaomi | Fine with the intension. Lenovo’s update seems more clear. |
| Ericsson | Support in general.In Rel-19, it was agreed that for option 4-1, the exchanged dataset can be associated with an ID for pairing. Hence, it shall be clarified that **a single pairing ID is associated with an exchange dataset.** Whether a pairing ID is associated with a single configuration or multiple configurations depends on how the dataset is structured. Based on the study from Rel-19, it is reasonable to assume that a single exchanged data set covers multiple configurations, meaning that one pairing ID is associated with multiple configurations.Is the third bullet intent to address whether quantization codebook/parameters may be different across different payload size configurations?The fourth bullet also requires further clarification. |
| MediaTek | Support |
| ZTE | OK with Lenovo’s rewording.For the first bullet, we think whether paring ID is PLMN unique shall be determined by RAN2.For the last bullet, we are not clear about the intention, please clarify on it whether the pairing ID is utilized for UE-side data collection? |
|  |  |

## 3.4 Quantization codebook

R18 study evaluated quantization aware and non-aware training. As quantization non-aware training suffers performance loss, the quantization aware training should be adopted.

### ***Proposal 4-1:***

***For Option 4-1 under Direction A in AI/ML based CSI compression, support quantization related information in the standardized dataset.***

* ***FFS: Quantization type: scaler or vector quantization***
* ***FFS: Quantization codebook related parameters and configuration***
* ***FFS: Common or different quantization codebook for CSI payload size***

Please provide your view below, particularly the reasoning for support option 1 or option 2:

|  |  |
| --- | --- |
| **Company** | **View** |
| OPPO | Whether the quantization related information is needed in Option 4-1 should be further discussed. In our view, quantization itself is a part of encoder, hence no need to indicate this information explicitly. |
| Lenovo | It depends if a single dataset can be used for development of different models having different quantization method or not.Also similar to the previous proposal, we suggest the following wording:***For Option 4-1 under Direction A in AI/ML based CSI compression, support exchange of quantization related information along with dataset exchange ~~in the standardized dataset~~.***  |
| Qualcomm | We think quantization related information is to be specified as part of the payload configuration, only the codebook is exchanged for 4-1. |
| Huawei, HiSilicon | Support |
| Xiaomi | The necessary to exchange quantization related information should be first discussed. If the quantization method or parameter is specified, it is not necessary to exchange such assistance information.  |
| Ericsson | Support in general. In addition, the same quantization method shall be used for both model training (i.e., the quantization configuration parameters exchanged from NW-side to UE-side) and model inference. |
| MediaTek | Support. We believe the quantization codebook/parameter needs to be exchanged |
| ZTE  | For the first bullet, there is a typo ‘scaler’, which should be ‘scalar’.This proposal is also applicable to Option 3a-1, and Option 3a-1 also involves the transmission of quantization related information along with the parameters.So, we suggest the following wording:***For Option 4-1/3a-1 under Direction A in AI/ML based CSI compression, support exchange of quantization related information along with dataset/encoder parameters exchange ~~in the standardized dataset.~~*** * ***FFS: Quantization type: ~~scaler~~ scalar or vector quantization***
* ***FFS: Quantization codebook related parameters and configuration***
* ***FFS: Common or different quantization codebook for CSI payload size***
 |

# Appendix 1: Company proposals

***Ericsson:***

[Observation 1 Target CSI in the format of Rel. 16 eType II with new parameters can provide over 95% overhead reduction comparing to Float32 format with performance gain of 0.7%~5.4% in terms of layer 1 SGCS over PC#8.](#_Toc206156357)

[Observation 2 Based on the above, the following benefits have been identified by using Rel-16 eType II with enhanced codebook parameters as the target CSI:](#_Toc206156358)

[ Improve the quality of the training dataset, in particular for higher layers (layer 2, 3, 4).](#_Toc206156359)

[ To not limit the performance of the AI-based CSI compression with the performance of the legacy mechanism (e.g., ParComb 8).](#_Toc206156360)

[ Open the possibility of having consistent dataset quality for layer 1 and layer 2 regardless of the configured rank.](#_Toc206156361)

[ Uplift the restrictions on the applicability of the Parameter Combination for a certain rank (some parameters are not defined for ParComb 7 and 8 for rank = 3 and rank = 4).](#_Toc206156362)

[ Lower complexity compared to raw-channel based eigenvector calculation.](#_Toc206156363)

[Observation 3 Performance target serves as a guidance for the UE-side on the achievable/expected performance during the encoder training phase.](#_Toc206156364)

[Observation 4 SGCS is invariant to absolute phases, which makes it a better performance target compared to NMSE, if the phase of ground-truth target and/or encoder input is not standardized.](#_Toc206156365)

[Observation 5 The performance of CSI compression can vary depending on network configurations.](#_Toc206156366)

Based on the discussion in the previous sections we propose the following:

[Proposal 1 Conclude that the dataset content for Direction A, sub-option 4-1 consists of at least the following:](#_Toc206156367)

[ {Target CSI, CSI feedback} samples](#_Toc206156368)

[ Performance targets, including the associated configuration and input data for evaluating the performance](#_Toc206156369)

[ Quantization codebook, including the associated configuration](#_Toc206156370)

[ Dataset ID](#_Toc206156371)

[Proposal 2 The same target CSI format is defined for NW-side data collection for training and for performance monitoring, for dataset exchange for inter-vendor training collaboration Direction A, sup-option 4-1, and CQI determination.](#_Toc206156372)

[Proposal 3 Support Rel. 16 eType II with new parameters as the Target CSI format.](#_Toc206156373)

[Proposal 4 For dataset exchange from NW-side to UE-side for Direction A sub-option 4-1, reuse the CSI feedback format for inference.](#_Toc206156374)

[Proposal 5 For the performance target sharing, support the end-to-end (encoder-decoder model pair) based performance target only.](#_Toc206156375)

[Proposal 6 For the end-to-end (encoder-decoder model pair) based performance target sharing, support only SGCS-based type of performance metric.](#_Toc206156376)

[Proposal 7 Support multiple SGCS statistics (e.g., SGCS values at X-percentiles) as the type of performance target instead of using only a single mean SGCS value across all samples.](#_Toc206156377)

[Proposal 8 For the case of a single dataset or model parameter set containing/supporting multiple configurations (payload sizes, number of layers, max rank values, subbands, etc.), multiple performance targets are exchanged for different configurations.](#_Toc206156378)

[Proposal 9 The testing dataset (input data for evaluating the performance) shall be exchanged from the NW-side, e.g., the last X percent of the exchanged dataset shall be used for performance evaluation.](#_Toc206156379)

[Proposal 10 Support SQ as the only quantization method that is used for quantizing the latent space for both model training and model inference.](#_Toc206156380)

[Proposal 11 The exchanged dataset for Direction A, sub-option 4-1 consists of multiple sub datasets corresponding to different configurations, a single dataset/pairing ID is assigned for the exchanged dataset.](#_Toc206156381)

***FutureWei:***

Proposal 1: Confirm that the dataset content for inter-vendor training collaboration Option 4-1 in CSI compression via two-sided model Case 0 includes at least the following:

* An ID that can be used to identify the dataset
* {Target CSI, CSI feedback} which corresponds to the input and output of the encoder for UE-side model training
* Performance target to help UE-side assess the encoder performance
* Quantization codebook associated with the exchanged dataset for UE-side to quantize the encoder output to generate the quantized CSI feedback

Proposal 2: For target CSI in the exchanged dataset for Option 4-1 in CSI compression via 2-sided model Case 0, use Rel-16 eType II code book with enhanced parameters to generate high-resolution samples for UE-side model training.

Proposal 3: For CSI feedback in the exchanged dataset for Option 4-1 in CSI compression via 2-sided model Case 0, adopt the format of CSI feedback that is generated before the quantization procedure to enable more flexibility in UE-side model training.

Proposal 4: For performance target in the exchanged dataset for Option 4-1 in CSI compression via 2-sided model Case 0, support including either SGCS or NMSE as the performance metric depending on the target CSI format in the exchanged dataset.

Proposal 5: Regarding the format of the performance target in the exchanged dataset for Option 4-1 in CSI compression via 2-sided model Case 0, support:

* Option 2: distribution of the performance target, e.g., SGCS / NMSE for 5, 10, 20, 30 percentiles, etc.

Proposal 6: Regarding performance target in the exchanged dataset for Option 4-1 in CSI compression via 2-sided model Case 0, support including only one performance target value (either SGCS or NMSE depending on target CSI format in the exchanged dataset) assuming a scalable model structure will be adopted at UE-side in developing the actual encoder.

***Spreadtrum, UNISOC***

Proposal 1: For CSI feedback related information, the CSI payload size and quantization configuration should be indicated.

* For vector quantization,
	+ NW needs to deliver the codebook or related parameter to UE
* For scalar quantization,
	+ The quantization granularity and range should be delivered from NW.

Proposal 2: An ID (e.g., dataset ID) associated with exchanged dataset should be exchanged from NW to UE.

Proposal 3: The average SGCS or average NMSE should be exchanged along with dataset from NW to UE.

* Support option 1: Average performance target, e.g. average SGCS and/or average NMSE.
* Multiple performance targets should be exchanged for different configurations.

***Huawei, HiSilicon***

Proposal 1: RAN1 focuses on the content/format of the data/dataset of Direction A Option 4-1 as SA WG study inputs before RAN#110.

Proposal 2: Regarding the data format/content information of Option 4-1, discuss at least the following aspects

* Format of the Target CSI
	+ Type of Target CSI, e.g., precoding matrix and channel matrix.
	+ Format of Target CSI, e.g., scalar quantization or eType II-like quantization.
		- eType II-like quantization is applicable regardless the Target CSI type is precoding matrix or channel matrix.
	+ Dimension of the Target CSI (Tx port number, layer/Rx antenna number, subband number, etc.).
* Format of the CSI feedback
	+ Dimension of output latent.
	+ Whether the CSI feedback is before quantization or after quantization.
	+ Quantization information.

Proposal 3: Regarding the dataset construction of Option 4-1, discuss at least the following aspects

* , wherein the Pairing IDs can be unique per operator.
* Number of data samples in the dataset.
* Dataset split/segmentation information.
* Association between Target CSI and CSI feedback.
* Scalability information. For different Tx port values, subband values, and CSI payload size values, separate data samples are provided, and their association is indicated.
* Performance target information
	+ Metric type, e.g., NMSE, MSE or SGCS.
	+ Metric statistic method, e.g., mean value and/or statistic values of X%CDF
	+ Performance target should be applied to at least Encoder and Decoder. End-to-End only is not considered.

***InterDigital***

Observation 1: Per Rel-19 discussion, target CSI consists of the precoding matrix may be represented in different domains, e.g., frequency, angular-delay domain, etc.

Proposal 1: For Target CSI in Option 4-1, specify the type (i.e., domain) and format of Target CSI.

Observation 2: Beam domain processing results in low-dimensional channel samples that may be compressed using low complexity Autoencoder models, resulting in reduction in both memory and computational requirements relative to spatial domain AE.

Proposal 2: Study beam domain processing for representing the Target CSI for Option 4-1.

Observation 3: For Option 4-1, scalability issues may arise when target CSI must accommodate varying input dimensionalities, as a change in the number of features necessitates transferring a new dataset.

Proposal 3: For Option 4-1, support dataset preprocessing as an additional information to be exchanged to handle the dataset scalability issue.

Observation 4: For Option 4-1, scalability issues may arise when CSI feedback transfer must accommodate varying input dimensionalities, as a change in the overhead size necessitates transferring a new CSI feedback.

Proposal 4: For Option 4-1, support structured CSI feedback with masking/truncation capabilities to cover a wide range of CSI feedback sizes with the one CSI feedback data.

Observation 5: The choice of E2E vs encoder only for the performance target additional information in Option 4-1 may impact which metric to use (SGCS, NMSE or both) as well as the contents of the additional information.

Proposal 5: RAN1 to determine whether the performance metric included as additional information along with the exchanged dataset in Option 4-1 refers to the E2E encoder-decoder model or to the encoder only.

Observation 6: For Option 4-1, it may be beneficial to include decoder backbone information, as well as the loss function definition along with the exchanged dataset.

Proposal 6: Study inclusion of the decoder backbone information and the loss function as additional information along with the exchanged dataset in Option 4-1.

***Google***

Proposal 1: The dataset for the inter-vendor training collaboration should include W2 and the compressed W2

* All the coefficients for W2 and compressed W2 are included

***CATT***

Proposal 1: For inter-vendor collaboration sub-option 4-1, spatial-frequency domain input can be prioritized for dataset format standardization.

Proposal 2: For format of target CSI in dataset for sub-option 4-1, the following options can be supported:

Option 1: Rel-19 eType II codebook (with or without enhancement)

Option 2: floating point

Proposal 3: For inter-vendor collaboration sub-option 4-1, both SGCS and NMSE can be supported as performance target.

SGCS can be used for UE-side training with nominal decoder

NMSE can be used for UE-side training without nominal decoder

Proposal 4: Regarding details of the format of the performance target, at least support Option 1.

Option 1: Average performance target, e.g. average SGCS and/or average NMSE

Proposal 5: For inter-vendor collaboration sub-option 4-1, support multiple performance targets corresponding to different configurations.

Proposal 6: For quantization alignment between UE-side and NW-side, prioritize uniform quantization.

Proposal 7: For the dataset content of inter-vendor collaboration sub-option 4-1, the additional information also includes:

* Pairing and/or associated ID
* Model structure related information
	+ Indicating specified model backbone type, as well as hyper parameters if needed
* Configurations related information
	+ E.g. layers, number of Tx ports, payload sizes, number of subbands

Proposal 8: Standardized dataset format should support different inference configurations at least regarding the following aspects:

Different number of Tx ports

Different subbands configurations

Different payload sizes

Proposal 9: The standardized dataset format should facilitate the aggregation of multiple configurations to reduce the dataset size.

Proposal 10: Method of realizing model scalability can be provided along with dataset for sub-option 4-1 as additional information.

***Vivo***

1. Both NMSE and SGCS can be used as performance target shared as additional information along with the exchanged dataset.
2. For performance target, there is one to one mapping relation between NMSE and SGCS. The NW can well map the performance target to either NMSE and SGCS without ambiguity.
3. The performance metric gaps between different ports or payloads are significant, whereas the gaps between different subbands are minimal.
4. If R16 eType II CB with legacy parameters (e.g., PC8) can be reused for a type of target CSI, the following codebook information shall be provided along with the Target CSI to enable recovery of the precoding matrix
	* Indicator of total number of non-zero coefficients
	* n1-n2
	* paramCombination(e.g., PC8)
	* subband number
	* R: numberOfPMI-SubbandsPerCQI-Subband
5. CSI feedback is Post-quantized CSI, which is a binary sequence with payload bits:{b\_1,b\_2,…,b\_payloads}, and payload information shall be indicated along with the CSI feedback.
6. For a data sample, the following mapping relationship between target CSI and CSI feedback can be considered
	* One-to-many mapping: one target CSI can be compressed to multiple CSI feedbacks. e.g., {target CSI, CSI feedback#1, , CSI feedback#i, CSI feedback#N}
7. The Pairing ID and associated ID (if necessary) should be exchanged along with the exchanged dataset.
8. Both SGCS and NMSE are supported as performance targets
	* Average SGCS and average NMSE can be considered
9. Multiple performance targets can be exchanged for different configuration (e.g., different ports and different payloads)
10. For the definition of SGCS:
	* For a given layer , subband , and data instance ，SGCS is defined as

where is the reconstructed precoder by NW-side decoder of *l-th* layer and n3-th subband, n4-th data instance, and is the corresponding precoder represented by PMI used for ground-truth target CSI for the *l-th* layer, n3-th subband, n4-th data instance.

* + And then average SGCS is calculated by
		- * + wideband frequency granularity
				+ multiple data instances in a dataset
				+ per layer
1. For the definition of NMSE, it is NMSE{CSI feedback, } for encoder-only performance target.
	* For a given layer , and data instance ，NMSE is defined as

where is the calculated CSI feedback based on UE-side CSI compression of *l*-th layer, n4-th data instance, and is the corresponding CSI feedback provided by NW for *l*-th layer, n4-th data instance .

* + Average NMSE is calculated by
		- * + multiple data instances in a dataset
				+ per layer
1. For a dataset content, the following can be included:
	* Pairing ID
	* performance target
	* Data sample number(N) / Data sample type (one-to-one, or one-to-many)
	* N Data sample is consisting of {target CSI, CSI feedback}
		+ - * Target CSI can be R16 eType II CB with legacy parameters (e.g., PC8), and codebook information can be associated with target CSI
				* CSI feedback is Post-quantized CSI, and payload information can be associated with CSI feedback

***Xiaomi***

Proposal 1: The exchanged data contents include target CSI, CSI feedback and additional information, where additional information may include multiple performance targets model pairing IDs and model scalability information.

Proposal 2: Support legacy eType II codebook quantization for target CSI.

Proposal 3: Support scalar quantization for CSI feedback.

Proposal 4: Support Option 1, i.e., average performance targe as a baseline for define the format of the performance target.

***TCL***

Proposal 1: Option 1, 2 should not be standardized, at least they are out of the scope of R19.

Proposal 2: RAN 1 should down select among option 3, 4 and 5 considering if unified model format or structure is shared between the NW and UE side model respectively.

* For option 4, there may be no need for offline-engineering.
* Option 3 and 5 are preferred if model structure or format is exchanged from NW to UE.

***ZTE***

General views

Proposal 1: The following specification impacts for the two sub-options of Direction A can be discussed with equal priority:

* For Direction A sub-option 4-1,
	+ Further consider dataset related information, e.g., dataset content, dataset type, dataset format, and additional information.
* For Direction A sub-option 3a-1,
	+ Further consider parameter related information, e.g., parameter content/format, quantization method and additional information.

Normative work on inter-vendor training collaboration

* Direction A sub-option 4-1

Proposal 2: For target CSI type for model training, support precoding matrix with spatial-frequency domain representation as a starting point.

Observation 1: Taking latent vector before quantization as CSI feedback in dataset is beneficial for improving UE-side encoder training performance. Nevertheless, taking bit sequence after quantization as CSI feedback can reduce exchanged overhead significantly.

Proposal 3: For CSI feedback in the dataset, the following two types can be further discussed:

* Latent vector before quantization
* Bit sequence after quantization

Proposal 4: For CSI feedback in the dataset, further discuss whether to use the same quantization codebook across varying payload size configurations.

Observation 2: If a dataset only corresponds to one configuration (e.g., Tx port, subband, payload size), the dataset ID is able to imply information about the configuration.

Observation 3: A dataset corresponding to multiple payload sizes with optimal format design can reduce overhead of dataset exchanging.

Proposal 5: Further discuss the necessity of including multiple configurations of target CSI (e.g., Tx Ports, subbands) in one dataset.

Proposal 6: For Direction A sub-option 4-1, further discuss SGCS or NMSE as the performance metric for performance target for potential down-selection.

Proposal 7: For Direction A sub-option 4-1, further discuss the necessity of associating each configuration combination of payload size, Tx port, and subband with a separate performance target.

Proposal 8: For Direction A sub-option 4-1, whether pairing ID should be global or local and other related issues shall be determined by RAN2.

Proposal 9: For Direction A sub-option 4-1, the necessity of transferring layer indication information (e.g., layer index for data sample) should be considered even for layer common model, given that each layer may have separate performance target.

* Direction A sub-option 3a-1

Proposal 10: For Direction A sub-option 3a-1, scalar quantization can be applied for the exchanged model parameters.

Proposal 11: For Direction A sub-option 3a-1, further discuss SGCS or NMSE as the performance metric for performance target for potential down-selection.

Proposal 12: For Direction A sub-option 3a-1, further discuss the necessity of associating each configuration combination of payload size, Tx port, and subband with a separate performance target.

Proposal 13: For Direction A sub-option 3a-1, whether pairing ID should be global or local and other related issues shall be determined by RAN2.

Observation 4: For scalability over the feature dimension, Alt 1 (specific embedding layer for each feature size) necessitates a specific embedding layer (e.g., a Fully-Connected layer) for each distinct feature size, leading to an increased number of model parameters and computational complexity.

Observation 5: For scalability over the feature dimension, Alt 2 (a common embedding layer with padding) employs padding operations to fulfil the predefined feature sizes without introducing extra complexity to either the model architecture or inference process.

Observation 6: For scalability over the token dimension, Alt 2 (padding at the input) is a simple way to operate and incurs minor performance degradation compared with the dedicated model design.

Observation 7: For scalability over payload configurations, compared with Alt 1 (specific output linear layer for each payload configuration), Alt 2 (truncation/masking of the output linear layer output) can reduce number of model parameters significantly.

Proposal 14: For model structure scalability for Direction A sub-option 3a-1, support

* For the choice of token dimension and feature dimension,
	+ Alt 1: Use subband as the token dimension and Tx port as a feature dimension
		- The number of tokens varies with the number of subbands.
* For scalability over the feature dimension,
	+ Alt 2: A common embedding layer with padding
* For scalability over the token dimension,
	+ Alt 2: Padding at the input
* For scalability over payload configurations,
	+ Alt 2: Truncation/masking of the output linear layer output
* Direction C

Proposal 15: For Direction C, support Transformer as the reference model backbone, avoiding the duplicated specification efforts between RAN1 and RAN4.

***Samsung***

Observation#1: For NW-side data collection, including in the inter-vendor training collaboration Direction A (Option 4-1), the UE logs and reports high-resolution Target CSI. A new format is expected be defined for Target CSI report.

Observation#2: For Direction A, NW-side may share the Associated ID(s) with dataset, i.e., {Target CSI, CSI-feedback}. UE-side may train its respective actual encoder based on the shared reference encoder/dataset and UE-side data (Target CSI) collected with the same Associated ID(s). Alignment on data categorization through the Associated ID enables training without NW-side and UE-side data distribution mismatch.

Observation#3: For inter-vendor collaboration Direction A Opt. 4-1, for inference, the UE reports CSI feedback based on a new format to be specified for the output of the encoder.

Observation#4: In dataset sharing based approach (Direction A Opt. 4-1), the high-resolution Target CSI is exchanged in two cases:

* Case1: In UE’s report of Target CSI for NW-side data collection
* Case2: In NW-side dataset {Target CSI, CSI feedback} sharing
* Note: Case 1 and Case 2 are expected to share the same high-resolution CSI format.

Observation#5: In dataset sharing based approach (Direction A Opt. 4-1), the CSI feedback is exchanged in two cases:

* Case1: In NW-side dataset {Target CSI, CSI feedback} sharing
* Case2: In UE’s CSI (inference) report
* Note: Case 1 and Case 2 are expected to share the same CSI feedback format.

Proposal#1: In dataset sharing based approach (Direction A Opt. 4-1), support a unified format for Target CSI and CSI feedback

* Target CSI format to be the same as the Target CSI format to be specified for NW-side data collection in AI 10.1.1.2
* CSI feedback format to be the same as the CSI feedback format to be specified for inference report in AI 10.1.1.1

Observation#6: For Option 4-1 of Direction A, performance degradation is observed when the assumed NW-side encoder backbone associated with the dataset is different from the UE-side encoder backbone.

Proposal#2: For Option 4-1 of Direction A, consider NW-side sharing the encoder backbone assumption associated with the dataset as additional information.

***NEC:***

Proposal 1: For inter-vendor collaboration under Direction A and sub option 3a-1, a standardized signalling mechanism can be defined to facilitate the transfer of AI/ML models or model parameters from the gNB (network side) to the UE (user equipment side) to support CSI compression.

Proposal 2: For inter-vendor collaboration under Direction A and sub option 3a-1, a standardized signalling mechanism can be defined from the UE to the gNB to indicate that the AI/ML model is ready for inference.

Proposal 3: For inter-vendor collaboration under Direction A and sub-option 3a-1, the potential specification impact of training should include considerations of model structure parameters, quantization configurations.

Proposal 4: For inter-vendor collaboration under Direction A and sub option 3a-1, NW should assign a unique identifier (e.g., model ID) to each transmitted encoder during encoder model (structure plus parameter) exchange.

Proposal 5: For inter-vendor collaboration under Direction A and sub option 3a-1, a standardized signalling mechanism can be defined from the gNB to the UE to initiate the inference procedure at the UE side.

Proposal 6: For inter-vendor collaboration under Direction A and sub option 3a-1, the gNB can indicate performance targets to the UE as part of the CSI feedback configuration.

Proposal 7: For inter-vendor collaboration under Direction A and sub option 4-1, a standardized signalling mechanism can be defined to facilitate the transfer of datasets from the gNB (network side) to the UE (user equipment side) to support AI/ML-based CSI compression.

Proposal 8: For inter-vendor collaboration under Direction A and sub option 4-1, a standardized signalling mechanism can be defined from the UE to the gNB to indicate the mapping between the dataset received and the trained AI/ML model—developed using that dataset—which is used for inference at the UE side.

Proposal 9: For inter-vendor collaboration under Direction A and sub option 4-1, NW should assign a unique identifier (e.g., dataset ID) to each transmitted dataset during encoder dataset exchange.

Proposal 10: For inter-vendor collaboration under Direction A and sub-option 4-1, the potential specification impact of training includes consideration of NW-side sharing of the encoder backbone.

Proposal 11: For inter-vendor collaboration under Direction A and sub option 4-1, a standardized signalling mechanism can be defined from the gNB to the UE to initiate the inference procedure at the UE side.

Proposal 12: For inter-vendor collaboration under Direction A and sub option 4-1, the gNB can indicate performance targets to the UE as part of the CSI feedback configuration.

Proposal 13: For inter-vendor collaboration under Direction C, a standardized signalling mechanism can be defined from the UE to the gNB to indicate the UE capability information, including details about supported standardized reference models for CSI compression.

Proposal 14: For inter-vendor collaboration under Direction C, a standardized signalling mechanism can be defined from the gNB to the UE to initiate the inference procedure at the UE side.

***Lenovo***

Observation 1: In option 4-1 and 3a-1 of Direction A, the performance of the two-sided model (especially if there exists UE data mismatch) is not clear if the UE directly train the encoder model (without the decoder model) based on the exchanged information.

Observation 2: In option 4-1 of Direction A, the UE can use the exchanged information to first train a nominal decoder and then uses the local decoder to train the encoder using the samples available at the UE-side. The trained encoder model can be used for new type UEs whose input data statistics have not been observed during the data collection/training step.

Proposal 1: For option 4-1 of Direction A, prioritize schemes based on first construction of the nominal decoder and then training of the encoder model.

Observation 3: For options 3a-1 and 4-1, associating an ID for pairing information during the exchange of dataset/model parameters between the UE-vendor and NW-vendor may disclose the information about the identity of the UE-vendor and/or NW-vendor during the inference phase.

Proposal 2: Consider specification of procedure/signaling enabling the UE/gNB to identify/select a correct paired encoder/decoder without disclosing identifying information of the other side, i.e., gNB/UE.

Observation 4: Direct exchange of dataset/model parameters between the UE-side and NW-side and associating ID for pairing discussion may lead to disclosure of the UE/NW vendor information during the inference phase. Dataset/model parameters exchange through a node/function in the network can potentially resolve this issue.

Proposal 3: To avoid disclosure of vendor-identity, specify required procedures/signalling to support exchange of dataset/model parameters through a node/function inside the core network managed by the OAM.

Proposal 4: Consider that the ID associated with dataset/model parameters is unique within the coverage area served by the node/function in the network responsible for dataset/model parameters exchange.

Proposal 5: Specify required procedures/signalling enabling the node/function (in the network responsible for dataset/model parameters exchange) to associate an ID with datasets/model parameters that need to be exchanged between the NW-side and the UE-side.

Proposal 6: Confirm that the ID associated with a dataset/model parameters can be used further to determine the pairing encoder/decoder.

Proposal 7: Further study the procedure to initiate development of an encoder model for a particular ID (i.e., dataset/model parameters) that the UE/UE-side has not yet developed the corresponding encoder model

Observation 5: Other than data transfer for model training, transmission of quantized ground-truth CSI might be required for sending channel information during model monitoring or model update phase. The acceptable data transfer overhead/latency for model monitoring/model update phases is usually lower than that of initial model training phase.

Observation 6: For fixed feedback overhead cases, there exist a trade-off between the number of samples that can be feedback and the resolution of each sample in the feedback data. Therefore, when analysing the gain of transmitting a dataset with higher sample resolution, it should be compared with the case of not increasing the sample resolution but instead send more samples.

Proposal 8: For transmission of ground-truth CSI samples, prioritize transmitting more samples rather than fewer samples with higher resolution per sample (e.g., more samples with current parameter configurations for Rel-16 Type II, instead of less samples with a new parameter configuration for Rel-16 Type II), especially for cases that the overhead is more important, e.g., ground-truth data transfer for model monitoring or model update.

Observation 7: Although it is desirable to have larger CSI dataset for training, due to the overhead associated with data collection, it is important to determine the proper size of the CSI dataset which provides enough information for training without imposing large overhead on the network.

Observation 8: To have a reduced training CSI dataset size, it is desirable to only include CSI samples which are more informative for model training.

Observation 9: Based on the statistics of the input data and the cost of data transfer, it might be beneficial not to transmit/transfer all CSI samples measured during the data collection phase. This might be due to lower information content of some CSI samples or high correlation between one CSI sample (a group of samples) and another CSI sample (group of samples).

Proposal 9: Support specification of procedures/signaling enabling transmission of subset of CSI samples among the set of measured/collected CSI samples from the environment.

Observation 10: Transmission of additional information such as the level of distortion when measuring a sample, quality indicator, or how frequent a particular sample (representative sample) occurs, can help the NW side to better train the model.

Observation 11: As the information content of CSI samples depends on the experienced distortion level, the UE may use distortion level or a quality indicator to determine which samples should be transmitted to the NW, or at least send this information along the sample itself. The NW may be able to use this information further during the training of the model.

Proposal 10: Support specification of procedures/signaling enabling transmission of subset of CSI samples based on the experienced distortion level or a quality indictor.

Proposal 11: Support specification of procedures/signaling for transmission of additional information such as sample-group size (how often samples are observed), quality indicator, distortion level along the transmission of the sample itself.

Observation 12: To reduce the UE data mismatch problem, the UE-side may use samples collected at the UE-side (along with the samples received from the NW-side) for training of the encoder model.

Observation 13: To consider statistics of the new UE-types in the design of the decoder model, the NW-side may collect new samples and use them (along with the original training dataset) to update/fine-tune the decoder model.

Proposal 12: Consider inclusion of the model ID in data collection configuration when collecting data for updating/fine-tuning of a certain model. The UE may also include the ID when reporting back the measured samples to the UE-side or NW.

***Panasonic***

Proposal 1: The specification works on reference model for Direction A sub-option 3a-1 and Direction C can be postponed waiting the conclusions of RAN4 study.

Proposal 2: RAN1 should start the discussion on dataset format based on the conclusion of target CSI type and data collection format.

Proposal 3: For performance target shared as additional information for inter-vendor training collaboration, SGCS is used for end-to-end training and NMSE is used for training encoder side.

Proposal 4: For performance target shared as additional information for inter-vendor training collaboration, multiple SGCS/NMSE statistics are supported.

Proposal 5: No model backbone / structure related information sharing between NW-side and UE-side could be sufficient.

***Oppo***

Proposal 1: The content of exchanged target CSI in Option 4-1 should be aligned with the content of target CSI in NW-side data collection.

Proposal 2: Only support CSI eigenvector in spatial and frequency domain as the target CSI content in Option 4-1.

Proposal 3: Regarding the format of target CSI in Option 4-1, different alternatives should be considered:

* Alt 1: float 32 target CSI format is used for NW-side data collection for model training:
	+ Alt 1a (baseline): float 32 target CSI format in Option 4-1
	+ Alt 1b: codebook-like based target CSI format in Option 4-1
* Alt 2: codebook-like based CSI format is used for NW-side data collection for model training:
	+ Alt 2a: float 32 target CSI format in Option 4-1
	+ Alt 2b: codebook-like based target CSI format in Option 4-1

Proposal 4: Suggest to study whether a unified format or multiple formats should be used for target CSI in Option 4-1, if different formats of target CSI are used in NW-side data collection for model training.

Proposal 5: Support bit output after quantization as the format of CSI feedback in Option 4-1.

Proposal 6: Support to use average SGCS instead of distribution of SGCS as the performance target for each layer in Option 4-1, including:

* Option A: direct SGCS
* Option B: differential SGCS compared to legacy codebook

Suggest to use Option B as the additional information.

***Nokia***

Proposal 1: RAN1 to continue the study of scalability for inter-vendor training Direction C following the existing work in RAN4.

Proposal 2: RAN1 to define the scalability and generalization parameters into three categories: external parameters (affect the format and size of the PMI data), internal parameters (affect the format and size of the compressed representation of the PMI), and implicit parameters (affect the statistical properties of PMI data).

Proposal 3: RAN1 to define the term *PMI interface description* to help with interoperability.

* Each *PMI interface description* has a single associated pairing ID that is unique (within appropriate scope, e.g. within the scope of pairing IDs managed by one network operator).

Proposal 4: RAN1 to consider the following configurations for scalability over internal parameters (overheads):

* Each PMI interface description should have a fixed (maximum) latent dimension and describe the unquantized output.
* The description may also describe a preferred quantizer and/or a preferred set of latent dimension truncation levels that are recommended to be used with this interface.
* Configuration of PMI feedback should include
	+ the ability to specify the quantizer to be used
	+ the ability to configure the use of a latent dimension smaller than the maximum latent dimension

Proposal 5: RAN1 to support configuration-specific models (sharing common core layers augmented by interface-specific adaptation layers) with the implementation scalability method.

Proposal 6: RAN1 to support scalable models (one model supports various PMI configurations) with the padding/interpolation scalability method.

Proposal 7: RAN1 to standardize dataset format that consists of entries, where each entry contains a structured set of information:

* Meta-information: includes essential details of channel conditions, such as scenario, UE-speed, carrier frequency, and antenna configurations.
* Target CSI/Precoding matrix: the uncompressed, ground-truth CSI (e.g., based on enhanced eTypeII codebook configurations), which serves as the training target.
* Paired Latent Vector (Direction A sub-option 4-1): The corresponding compressed latent vector before quantization.
* Reference Encoder (Direction A sub-option 3a-1 and Direction C): Partly specified reference encoder model (weights and structure) for Direction A sub-option 3a-1 and fully specified reference encoder model for Direction C.

Proposal 8: Standardization should cover data pre-processing, ordering convention to use for mapping PMI coefficients, and quantization schemes. Defining common normalization techniques (e.g., scaling and phase normalization of channel coefficients), ordering convention of PMI coefficients, and quantization schemes could ensure consistent model training and behavior across vendors.

Proposal 9: The dataset format and its contents should be linkable to the pairing ID.

***NTU***

Proposal 1: Consider the following steps for deriving the reference models (for encoder and decoder) for dataset generation:

* Step 1: Determine simulation setup and encoder input dataset generation procedure
* Step 2: Determine model type and parameter
* Step 3: Determine training hyper parameters
* Step 4: Determine the test decoder(s) by first determine evaluation methodology and selection/merge criterion, and then decide the test decoder(s) based on companies’ evaluation

Proposal 2: Consider the following guidelines based on R19 RAN4 discussions

* For simulation setup:
	+ First decide whether to use system level or link level simulation for determining test decoder
	+ Consider setup in RAN4 R19 TR as starting points, what are the parameters need to be updated
* RAN4 concluded that aggregated encoder input dataset is recommended for reference decoder evaluation since the performance variation across encoder/decoder pairs from different companies when input dataset from only one company is used
* RAN4 has selected MLP and CNN in R19 discussion, and these models can be the starting point for RAN1 discussion
* For training, RAN1 needs to decide whether to agree training related parameters, and if yes, what parameters to be agreed
* Evaluation method and criterion: whether to consider one more factors listed in the following
	+ Achievable performance metrics: SGCS or NMSE
	+ Complexity: including flops or model storage size, can consider to set an upper bound
	+ Robustness (based on performance metrics): whether the decoder can achieve satisfactory performance when connecting to encoder with different structures with the decoder, but trained with the decoder’s input and output dataset

***LG Electronics***

Observation #1: Most of issues for inter-vendor training collaboration can be led by other working groups, i.e., RAN4, RAN2 and SA.

Proposal #1: For Direction C, RAN1 can study on whether specification support for model scalability is needed.

Proposal #2: For sub-option 3a-1 in Direction A, RAN1 can study on at least following aspects: format of model parameter range, size of model parameters, and necessity of model parameter ID.

Proposal #3: For sub-option 4-1 in Direction A, RAN1 can study on at least following aspects: dataset format, the number of data samples in datasets, size of each component in dataset, and necessity of dataset ID.

***Apple***

Proposal 1: For inter-vendor training collaboration option 4-1, the dataset exchange should support scalable model training across number of subbands, antenna ports and payload size.

Proposal 2: For inter-vendor training collaboration Option 4-1, the additional information include performance targets defined in terms of SGCS and NMSE metrics:

* SGCS is used for Alternative 1 training, where the UE side first trains a nominal decoder.
* NMSE is used when the UE side directly trains the actual encoder, with NMSE calculated based on floating-point values before quantization.

Proposal 3: To enable quantization aware training when the quantization method is not explicitly specified, codebook information should be included as part of the dataset exchange.

Proposal 4: For inter-vendor training collaboration under Option 4-1, the dataset ID should be generated by the training entity. The dataset ID is PLMN unique, supports augmenting existing datasets to enable fine-tuning of the AI model and facilitates scalable AI training.

***Fujitsu***

Observation-1: The target CSI of Option 4-1 is obtained from UE’s CSI feedback.

Proposal-1: Regarding standardized dataset format/content for inter-vendor training collaboration, discussion on Direction A sub-option 4-1 is prioritized. While the details of model parameters in sub-option 3a-1 can be left for RAN4 study.

Proposal-2: The format of target CSI and the format of CSI feedback of Option 4-1 is suggested as:

* Target CSI: Quantization with eT2-like high-resolution codebook
* CSI feedback: 0/1-bit sequence

Proposal-3: Support precoding matrix over spatial-frequency domain as the data type of Direction A, sub-option 4-1.

Proposal-4: For Direction A, besides performance target, its associated test dataset should be shared from NW to UE.

* Specification support of test dataset/test target exchange over the air is suggested to be studied

Proposal-5: For Direction A, rank indication information and layer index information can be taken as additional information and shared from NW to UE along with the exchanged dataset/model parameters.

Proposal-6: For Direction A, besides the end-to-end performance target of NMSE and SGCS, support performance metric for UE-side encoder, and relevant performance target should be shared from NW to UE as the additional information.

Proposal-7: For Direction A, the exchanged dataset or the model parameters is associated with an ID. As an additional information, this ID can be used to facilitate alignment and pairing operations between the model parts of two sides.

Proposal-8: For Direction A, as additional information, hyper-parameters (e.g. segment size) and/or trainable parameters (e.g., VQ table) for quantization book can be shared from NW to UE along with the exchanged dataset/model parameters:

* FFS: Standardized configuration of the quantization codebook, and its scalability across payload sizes

***Sharp***

Proposal 1: For Direction A sub-option 3a-1, if the parameters exchanged from gNB are not available, UE performs CSI feedback (compression) with the specified model provided by RAN4.

Proposal 2: For Direction A sub-option 3a-1, the parameters exchanged at least consist of Tx ports, CSI feedback payload sizes, bandwidths, number of slots, and performance target.

Proposal 3: For Direction A sub-option 3a-1, UE directly trains the encoder with the parameters exchanged, i.e. Alt.2.

Proposal 4: For Direction A sub-option 4-1, adopt Alt.1 for generation model training, i.e. UE firstly trains a nominal reconstruction model and then trains a new generation model.

***MediaTek***

Discuss inter-vendor collaboration schemes which jointly use Direction A and Direction C.

For Direction C and Direction A with sub-options 3a-1 and 4-1, AI/ML model ID is needed for monitoring configuration.

For Direction C, AI/ML model and its ID can be specified together.

For inter-vendor collaboration option 3a-1 in Direction A, if the performance target is shared in form of NMSE and/or SGCS, consider: i) Sharing distribution of the performance target and ii) One distribution of performance target per configuration.

For Direction A with sub-option 3a-1, UE’s re-engineered encoder, NW’s decoder as well as the shared quantization codebook all together inherit the ID assigned to the shared parameters by NW.

For inter-vendor collaboration option 4-1 in Direction A, consider:

* Sharing NMSE/SGCS performance target in either average or distribution form
* Sharing backbone information NW’s encoder.

For Direction A with sub-option 4-1, UE’s encoder, NW’s decoder as well as the shared quantization codebook all together inherit the ID assigned to the shared dataset by NW as AI/ML model ID.

For Direction A with sub-option 3a-1 and sub-option 4-1, discuss whether and how to keep uniqueness of AI/ML model IDs across different NW vendors.

In Direction A with sub-option 4-1 for {Target CSI, CSI feedback} dataset exchange, consider:

* For Target CSI part, reuse the same format leveraged for NW-side data collection.
* For CSI feedback part, share output of NW’s encoder before quantization in float32 format.

***ETRI***

Proposal 1: For AI/ML-based CSI feedback, for inter-vendor training collaboration direction C sub-option 4-1, consider following information as additional information for training dataset:

* Performance target,
* Information on the backbone network, and
* Associated ID for consistency.

Proposal 2: For AI/ML-based CSI feedback, for inter-vendor training collaboration direction C sub-option 4-1, consider OTA based standardized signaling, or non-OTA based standardized signaling by OTA based control signaling.

Proposal 3: For AI/ML-based CSI feedback, for inter-vendor training collaboration direction C sub-option 3a-1, consider following information as additional information for training dataset:

* Performance target,
* Quantization information, and
* Associated ID for consistency.

Proposal 4: For AI/ML-based CSI feedback, for inter-vendor training collaboration direction C sub-option 3a-1, consider OTA based standardized signaling, or non-OTA based standardized signaling by OTA based control signaling.

***CMCC***

Proposal 1: For Option 4-1 in Direction A in AI/ML based CSI compression, support {target CSI, CSI feedback} as the standardized dataset content.

Proposal 2: At least precoding matrix in the spatial-frequency domain in a subband granularity should be supported as the standardized dataset content of target CSI for inter-vendor collaboration training Option 4-1.

Proposal 3: There should be a common design for target CSI for inference, monitoring and inter-vendor collaboration training.

Proposal 4: In CSI compression using two-sided model use case, regarding the standardized dataset format of target CSI for inter-vendor collaboration training Option 4-1, codebook-based Rel-16 eType2 and Rel-18 eType2 for PMI prediction can be supported.

Proposal 5: In CSI compression using two-sided model use case, regarding the standardized dataset format of target CSI for inter-vendor collaboration training Option 4-1, the basic codebook structure could be reused, along with the basic concept of spatial domain, frequency domain and Doppler domain basis.

Proposal 6: In CSI compression using two-sided model use case, regarding the standardized dataset format of target CSI for inter-vendor collaboration training Option 4-1, the exact supported values of codebook parameters can be studied to make sure high resolution data transfer.

Proposal 7: Some necessary model related information, such as model backbone, can be aligned between NW side and UE side to achieve better performance for Option 4-1.

Proposal 8: For inter-vendor-collaboration Options 4-1 in Direction A, for the type of performance metric, prioritize SGCS, and the definition of SGCS used for performance metric for AI/ML based CSI prediction can be reused as much as possible.

***Sony***

[Proposal 1: Adding additional information, which indicates training dataset or testing dataset, to assist CSI model training collaboration.](#_Toc194047595)

Proposal 2: Define standardized additional information to be exchanged between the NW and UE to ensure consistent training and operation of two-sided CSI models.

Proposal 3: model updates should be followed by coordinated parameter-related information exchange between NW and UE to maintain model alignment.

Proposal 4: model update triggering mechanisms (periodic, aperiodic, semi-static) need to be further studied for flexibility and efficiency.

***Qualcomm***

Proposal 1: For inter-vendor collaboration sub-option 4-1, each exchanged dataset should include the following:

* *Number of samples K*
* *One pairing ID #n*
* *Associated quantization codebook per payload configuration*
* *Multiple sets of samples,*
	+ *each set with a specific subband, port and payload configuration,*
	+ *each set contains K samples, each sample is pair-wise encoder input and output*

Proposal 2: For inter-vendor collaboration sub-option 4-1, adopt element-wise quantization using floating point for the samples of exchanged dataset.

Proposal 3: Common codebook per pairing ID is preferred. The necessity of specific codebook per subband configuration, per port configuration and per layer needs clear justification.

Proposal 4: To align the model design aspects, consider either of the following options:

* *Alt1: NW exchange tokenization and scalability options used in reference encoder input/output generation.*
* *Alt2: NW adopts the tokenization and scalability options of the RAN4 defined / specified model when developing their own reference encoder and generating the input/output.*

Proposal 5: For inter-vendor collaboration sub-option 4-1, NW provides both NMSE and SGCS target to the UE. Which performance target or both to use is upto UE implementation, and UE is expected to satisfy at least one of these two targets.

Proposal 6: Performance targets statistics should be exchanged, e.g., NMSE / SGCS at x% of the performance resulted by the exchanged dataset, or the NMSE / SGCS for each exchanged data sample.

Proposal 7: Specific performance targets are needed for particular layers, and are needed per subband / port / payload configuration.

***NTT Docomo***

Proposal 1

* RAN1 should start specifying the format for the model parameter exchange and tightly cooperate with RAN4.

Proposal 2

* For Option 3a-1, support using the model parameters of Option 1 reference models as a baseline for the model parameter exchange.
* Support the partial model parameter exchange and update on top of baseline parameters.

Proposal 3

* The specification works on the dataset exchange format can be postponed based on the conclusions of the SA2 study.

# Reference

R1-2506206 Work plan for NR\_AIML\_air\_Ph2, Qualcomm

R1-2505135 Inter-vendor training collaboration for two-sided AI/ML models Ericsson

R1-2505150 Discussion on inter-vendor training collaboration for two-sided AI/ML models FUTUREWEI

R1-2505163 Discussion on Inter-vendor training collaboration for two-sided AI/ML models Spreadtrum, UNISOC

R1-2505201 Inter-vendor training collaboration for two-sided AI/ML models Huawei, HiSilicon

R1-2505241 AI/ML CSI Spatial/Frequency Compression: Inter-vendor Collaboration InterDigital, Inc.

R1-2505262 Inter-Vendor Collaboration for AI/ML based CSI Compression Google

R1-2505301 Discussion on Inter-vendor training collaboration for CSI compression CATT

R1-2505407 Discussion on inter-vendor training collaboration for two-sided AI/ML models vivo

R1-2505454 Discussion on inter-vendor training collaboration for two-sided AI/ML models Xiaomi

R1-2505479 Discussion on inter-vendor collaboration for CSI compression TCL

R1-2505493 Discussion on inter-vendor training collaboration for two-sided AI/ML models ZTE Corporation, Sanechips

R1-2505575 Views on inter-vendor training collaboration for two-sided AI/ML models Samsung

R1-2505642 Discussion on Inter-vendor Training Collaboration for AI/ML models NEC

R1-2505689 Inter-vendor training collaboration for two-sided AI/ML models Lenovo

R1-2505701 Discussion on inter-vendor training collaboration for two-sided AI/ML models Panasonic

R1-2505748 Inter-vendor training collaboration for AI/ML CSI compression OPPO

R1-2505804 Inter-vendor training collaboration for two-sided AI/ML models Nokia

R1-2505806 Reference Model for Data Generation NTU

R1-2505821 Discussion on Inter-vendor training collaboration for two-sided models LG Electronics

R1-2505904 Inter-vendor training collaboration for two sided AI/ML models Apple

R1-2505964 Discussion on inter-vendor training collaboration for two-sided AI/ML models Fujitsu

R1-2506012 Discussions on Inter-vendor training collaboration Sharp

R1-2506031 Inter-vendor training collaboration for two-sided AI/ML models MediaTek Inc.

R1-2506059 Discussion on inter vendor training collaboration for two-sided AI/ML models ETRI

R1-2506090 Discussion on inter-vendor training collaboration for CSI compression CMCC

R1-2506109 Discussion on Inter-Vendor Training collaboration for Two-Side AI/ML Models Sony

R1-2506209 Inter-vendor training collaboration for two-sided CSI compression use case Qualcomm Incorporated

R1-2506296 Discussion on the inter-vendor training collaborations for two-sided AI/ML models NTT DOCOMO, INC.