**3GPP TSG-RAN WG4 Meeting #111 R4-240xxxx**

**Fukuoka city, Japan , May 20th – 24th, 2024**

**Agenda item:** 10.11.5

**Source:** Qualcomm Incorporated

**Title:** AI/ML ad-hoc meeting minutes

**Document for:** Approval

# Introduction

The discussion on the AI/ML study is organized under a single thread [133] in RAN4#111. The ad-hoc meeting will discuss some of the topics from the moderator summary in [1].

# Discussion

## Testability and interoperability issues for beam management

### Sub-topic 2-5

*Test setup*

A list of test setup/needs should be created in order to see what kind of test setup is needed and what is feasible

**Issue 2-5: Test setup needs**

* Proposals
  + Option 1:
    - number of Tx beams
    - AoA, AoD
    - Propagation conditions (including need for LoS/NLoS)
    - UE rotation yes/no
  + Option 2:
    - other parameters
* Recommended WF
  + To be discussed

Discussion:

Apple: [2-16] set B Tx beams

Nokia: 16 is the max number of beams for set B

Samsung: 2 is too little . for set A, gNBs support about 64 beams. 1/4 of the overhead means we would need 16 beams

Nokia: how dynamic, it is difficult to say. it would be needed anyway

Samsung: for spatial prediction case 1, is rotation really needed? if we add some rotation we can test the Rx codebook impact. do we need to make the system so complex. vibration during rotation needs to be considered, might impact the results. if we can test in a static way, it’s a better start.

Qualcomm: including channel model, Tx codebook and UE rotation, we would talk about the Rx RSRP distribution. these would have influence on the distribution seen. that would introduce a linear transformation. by rotation we would test impact of different RSRP distribution on the model. we can focus on the basic setup. once we have the basics.

Keysight: UE rotation is too slow to introduce dynamic change of channel conditions. we need a positioning system but no rotation during a test

R&S: Ue rotation is a metric to emulate real environment when the user is moving. by rotating we could emulate a very precise trajectory.

Samsung: UE rotation make us simulate UE rotation, not moving.

KTL: for spatial characteristics, static is ok. otherwise, we need to combine rotation with a dynamic channel for temporal prediction

Apple: for set A, we might need [64]

Samsung: in some small cells, we might have less than 64

Moderator: we need a max value

Agreement:

maximum number of set B Tx beams that test system should be able to emulate: [8-16]

maximum number of set A Tx beams that test system should be able to emulate: [64-128]

FFS on AoAs

UE rotation during the test: FFS

UE rotation/repositioning between different tests: Yes

### Sub-topic 2-6

*Data sets*

**Issue 2-6: Datasets for training/testing**

* Proposals
  + Option 1: Training Data set to be specified in RAN4(directly or through some algorithm )
  + Option 2: Training data set to be left to implementation (companies can generate it based on knowledge of the test environment)
  + Option 3: others
* Recommended WF
  + To be discussed

Discussion:

Qualcomm: Option 2 is the Tx codebook is specified. if we have the channel model, we should have enough knowledge to generate training data

Samsung: Option 2 is more reasonable. Rx codebook is up to UE implementation, UE vendors would take that into account when training the UEs

Docomo: dataset for testing should be specified. what data should be used for test. dataset for training is up to implementation but dataset for testing is specified.

Ericsson: we do not think we need to say that it is based on test environment.

Xiaomi: does the test environment/condition mean also channel model

Moderator:yes

Nokia: will the model in the UE be trained for the test or for a real deployment

Moderator: the only way to do that is to make the test as realistic as possible.

Qualcomm: beam prediction test, the model highly depends on the codebook. unless we have a model from the infra vendors, there is no way to guarantee that UE will perform in the field.

Ericsson: we sill have concern

Mediatek: we just show that performance with different codebooks is different, UE needs to know the Tx codebook to have good performance.

Apple: question to Ericsson and Nokia. what is your concern? between Option 1 and 2 or you have another option? infra-vendors can provide codebooks for that. the current tests, they are all simplified. why is now so different?

Samsung: we are discussing overfitting and generalization, we can discuss generalization over different test environments. if the network vendors want to use a different codebood in the field, this is not possible in 3GPP testing.

Ericsson: we should just say training data is left to implementation

Moderators: infra vendors have concerns that UEs model could overfit to the test environment and have poor field performance

Qualcomm: we can agree to this

Agreement:

* Vendors may take into account the test environment/conditions defined by RAN4 when training the UE
  + sufficient test environment/conditions should be defined to enable vendors to create the data needed for training
* FFS on proposals to augment training data to avoid overfitting of UE models to the test environment
* RAN4 to strive to make the test conditions similar to field deployment conditions

### Sub-topic 2-7

*Beamforming consistency*

multiple companies brought up the need to discuss the consistency between set A and set B, otherwise it is not expected that beam prediction would work

**Issue 2-7: Consistency**

* Proposals
  + Option 1: consistency between set A and set B is reflected by below aspects:
    - Same NW antenna/beam configurations for set A and set B, and that its configurations don’t change during training and inference.
    - Same channel model for set A and set B, and that its configurations don’t change during training and inference. (Question: if consistency is valid for a non-static(time-varying) channel for set A and set B?)
  + Option 2: Consistency should be defined in a different way
  + Option 3: no need for any consistency definition
* Recommended WF
  + To be discussed

Discussion:

Apple: what does the 2nd subbullet mean? same channel model. same model for training and inference.

Samsung: consistency in RAN1 is between training and inference. for the test, should be straightforward because it’s obvious that this spatial consistency is needed.

Qualcomm: first bullet says set A and set B doesn’t change during inference and training. we do not need the first sentence.

Intel: based on the discussion, we would not discuss consistency between training and inference.

### Sub-topic 2-8

*Measurement error impact*

Multiple companies brought up the need to evaluate the impact that the UE measurement error and the error in the training data have on inference accuracy

**Issue 2-8: Measurement error impact**

* Proposals
  + Option 1: RAN4 should study the impact of measurement error, companies should bring proposals for the next meeting on how to proceed with such a study
  + Option 2: RAN4 should postpone the discussion on the impact of the measurement error for now
  + Option 3: No need to evaluate the measurement error impact
* Recommended WF
  + To be discussed

### Sub-topic 2-9

*UE reporting for network side models*

One company brought up a possible need for RAN4 to introduce different reporting schemes to help train the network side models.

**Issue 2-9: UE reporting for network side models**

* Proposals
  + Option 1: RAN4 shall introduce the necessary core requirement on supporting data collection for NW-side AI/ML model inference/training (for BM-Case1 & 2), by considering:
    - Potential enhancement on L1 measurement/reporting for inference: e.g., beam reporting for more than 4 beam in L1 signaling, and overhead reduction;
    - Potential enhancement on MDT-based measurement/reporting for training.
  + Option 2: RAN4 cannot introduce new reporting schemes, proposal should be made in another WG
  + Option 3: others
* Recommended WF
  + To be discussed

## Testability and interoperability issues for positioning accuracy enhancement

### Sub-topic 3-1

*Requirements for case 1*

In the previous meeting it was agreed to postpone the discussion until a reporting scheme, if defined, is clear. Some companies are proposing not to define any requirements for this case

**Issue 3-1: Requirements for case 1**

* Proposals
  + Option 1: RAN4 should not define requirements for case 1
  + Option 2: RAN4 should postpone the discussion until other WG conclude on defining a reporting scheme or not
  + Option 3: Others
* Recommended WF

To be discussed

Note: currently there are no requirements for UE based positioning

Discussion:

Docomo: we have strong concern for option 1, we need some requirement.

Moderator: please bring some analysis on how it could be defined

CMCC: we also have concern on option 1, in legacy positioning, UE need to report metrics to LMF, we could maybe use something similar. we can wait to see if other groups define some reporting metrics.

vivo: agree with CMCC. we do not need to preclude having requirement. RAN1 has several candidates to report

Nokia: we have concern on option 1. UE can report even coordinates.

Ericsson: RAN2 should define this first. we shouldn’t talk about requirements yet. we are talking about performance requirements, not monitoring.

Apple: we agree with Ericsson. privacy from UE side is an issue. it is not feasible to define this

### Sub-topic 3-2

*Requirements for case 2a*

Requirements for case 2a have been deprioritized, however, some companies are proposing to discuss this anyway.

**Issue 3-2: Requirements for case 2a**

* Proposals
  + Option 1: RAN4 should not define any positioning accuracy requirements because positioning is LMF based
  + Option 2: RAN4 should continue to discuss how to define requirements for case 2a, companies should bring more concrete proposals in future meetings
  + Option 3: Postpone discussion until other groups make more progress
  + Option 4: others
* Recommended WF
  + To be discussed, options are not exclusive

Discussion:

vivo: option 3 is better. RAN4 must define requirement for other

Nokia: we want to add a comment about intermediate requirements. LOS/NLOS could have a requirement.

Agreement:

No requirements on LMF for positioning accuracy

FFS on RAN4 requirements for any UE reported measurements defined by other groups

### Sub-topic 3-3

*Requirements for case 2b*

Requirements for case 2b have been deprioritized, however, some companies are proposing to discuss this anyway.

**Issue 3-2: Requirements for case 2b**

* Proposals
  + Option 1: RAN4 should not define any positioning accuracy requirements because positioning is LMF based
  + Option 2: RAN4 should continue to discuss how to define requirements for case 2b, companies should bring more concrete proposals in future meetings
  + Option 3: Postpone discussion until other groups make more progress
  + Option 4: others
* Recommended WF
  + To be discussed, options are not exclusive

Agreement:

No requirements on LMF for positioning accuracy

FFS on RAN4 requirements for any UE reported measurements defined by other groups

### Sub-topic 3-4

*Requirements for other reported metrics*

Some companies are proposing to already start discussing how to define requirements for possible reported metrics such ToA, LoS/NLoS, etc. These might be introduced for cases 2a/2b, 3a/3b

**Issue 3-4: Requirements for reported metrics**

* Proposals
  + Option 1: RAN4 to postpone discussion until reported metrics become clear in other groups
  + Option 2: RAN4 to already start the discussion on how to define requirements for LoS/NLoS indicator
  + Option 3: RAN4 to start discussing how to define requirements for other reported values, e.g. CIR/PDP, ToA,
  + Option 4: others
* Recommended WF
  + To be discussed, options are not necessarily exclusive

## Testability and interoperability issues for CSI compression and CSI prediction

### Sub-topic 4-1

*Reference encoder/decoder definition*

Definition of reference encoder/decoder was discussed in the previous meeting but was not agreed.

**Issue 4-1: Reference encoder/decoder**

* Proposals
  + Option 1:
    - the encoder/decoder used in RAN4 discussions at least for simulation alignment/requirement derivation, test decoder derivation and/or test decoder verification. It could be documented (in TR, WF, etc) or captured in the specifications as necessary.Option 2: Other definitions
  + Option 2:
    - Reference decoder/encoder: The decoder/encoder model used to define the minimum performance requirements. The Reference decoder is identical to the Test decoder.
  + Option 3: other definition
* Recommended WF
  + Option 1

To be discussed if any clarifications are needed

### Sub-topic 4-1bis

*CSI-compression scheme*

The scheme of AI/ML-based compressed CSI feedback for parameters and metric alignment.

**Issue 4-1bis: AI/ML based CSI compression scheme**

* Proposals
  + Option 1: Agree on the baseline scheme for CSI compression use-case.
  + Option 2: other
* Recommended WF

To be discussed if any clarifications are needed.

### Sub-topic 4-2

*Standardization steps for Option 3*

A flow chart for the work on Option 3 was presented in the previous meeting (R4-2405653, reproduced below) and discussed. This was discussed, however there was no formal agreement. A refinement to this chart was proposed in R4-2407236. Also, some proposals were made in R4-2407368. These should be discussed to further the align the steps needed to progress on Option 3 feasibility.

**Step-1: Identify necessary Model Architecture Parameters**

**Standardization Procedure End**

for a certain use case  
(e.g., CSI compression for precoding matrix under certain config.)

Model architecture parameters could include: Model type, Model depth, Layer type/size, Quantization, etc.

Model training procedure, loss function, training datasets, hyperparameters, etc.

**Step-3: Companies provide two-sided model design based on their own study/preference**

**Step-4: Performance comparison based on different companies’ en/decoder designs**

**Yes**

**No**

**Step-6: Performance alignment by companies based on agreed model architecture/training parameters**

Performance comparison in terms of metrics like NMSE, SGCS, etc.

**No**

**Yes**

**No**

**Standardization   
Procedure Start**

**Step-2: Identify necessary Model training Parameters**

Test decoder is expected to be generated in this step

Reference encoder is assumed, but leave enough implementation flexibility to vendors (similar to Demod alignment for MMSE-IRC)

**Step-8: performance alignment   
for encoder design by companies   
based on assumptions on reference encoder**

**Yes**

RAN4 performance requirement obtained (for certain reference encoder)

**Step-10: Derive RAN4 performance requirement**

**Step-5: RAN4  
agree on two-sided model architecture  
 / training parameters?**

**Step-7: RAN4 agree on test decoder   
(which can be fully specified in spec.)**

**Step-9: RAN4 achieve performance alignment?**

**Issue 4-5: Option 3 standardization process**

* Proposals
  + Option 1: Flowchart below better represents the steps needed in RAN4, should be considered as baseline:

**A diagram of a process

Description automatically generated**

* + Option 2: Following changes should be made to the chart from R4-2405653(discussed in RAN4#110Bis)
    - Add a step, “identify target cases with specific test conditions” before step 1.
    - For step 1, suggest to limit the scope of potential model types. Transformer-based could be prioritized.
    - Add a step, “Define evaluation methodology” before step 4.
    - Suggest to align the understanding of step 5 on, whether only a single model architecture would be determined for 2-sided case at least for this release.
    - Discuss the necessity of step 8 and 9.
  + Option 3: other changes
* Recommended WF

To be discussed

### Sub-topic 4-3

*Decoder parameters for Option 3*

Several companies made proposals for information that should be agreed in order to be able to derive a full decoder for Option 3. Some agreements on needed parameters were reached in previous meetings and documented.

**Issue 4-3: Decoder parameters**

* Proposals

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Vivo (updated) | CATT | QC (updated) | Xiaomi | Intel | E/// (updated) | Apple (updated) | Nokia (updated) | Keysight | ZTE (updated) | Samsung (updated) |
| Model type | Transformer, CNN, RNN, MLP | Transformer | MLP | Transformer | Several model types can be considered (e.g., transformer, CNN) RAN1 inputs on the best identified models in terms of performance/complexity can be requested | Transformer or CNN depending on design target | Transformer, CNN, RNN, MLP | Transformer |  | Transformer | Transformer |
| Model depth | Number of layers |  | Three linear layers (with one activation function) | 6 | decide upper bound of complexity based on RAN1 evaluation |  | Number of layers, CNN: Kernel/Filter Size, Padding, Stride, Pooling layers parameters, Number of channels | Several multi-head attention layers (min: [3], max: [7]) |  |  | 4 layer |
| Layer type | Fully connected, convolutional, activation layer, normalization layers, etc. |  | MLP with expansion factor N = 4, and each layer/function is described in the following • 1st linear layer: input is latent message of size Zdim and output is a vector of size nSB x nTx • Reshape: convert the vector of size nSB x nTx to nSB vectors with size nTx • 2nd Linear layer: For each subband, the input is a vector of size nTx, and the output is a vector of size N x nTx. The same linear layer is applied to each of nSB subbands. • Activation function: GELU • 3rd Linear layer: the input is a vector of size N x nTx, and the output is with a vector of size nTx. The same linear layer is applied to each of nSB subbands. | Scalar |  |  | Fully connected, convolutional, activation layer (activation type: leakyRelu,etc), batch(group)-normalization layer,dropout layer, etc. | Fully connected layers with activation function for each attention layer/block. |  |  | Fully connected layers with activation function for each attention layer/block |
| Layer size | Neuron count and configuration |  |  |  |  |  | Neuron count and configuration | *Note that output layer can be different.* |  |  | Scalar quantizer, 2 bits per dimension |
| Quantization method for the encoder output | Scalar, vector (with codebook) |  | scalar quantizer, 2 bits per dimension (element) | int16 |  |  | Scalar, vector (with codebook) | Specify embedding and feedforward dimensions, number of attention heads per attention layer/block. |  |  | 142bits (for 2-layer case) |
| Encoder-decoder interface | Number of bits of latent message |  | Use power of 2, choose from 32, 64, 128, 256 and 512 bits. | 60, 120, 280bits |  | Consider 63, 110 or 230 | Number of bits of latent message | FFS, e.g., 64 latent dimensions with two-bit quantization, i.e., 128 overhead bits. |  | 57，104，270bit | 86bit (for 1-layer case) |
| Fixed point representation | Int8, int16, floating point etc |  |  |  |  | Int8, int16, floating point etc. | Int8, int16, floating point etc | FFS, decision to be made during/after model design, or may be left for implementation. |  |  | N/A |
| Format of input to encoder/output of decoder |  |  |  |  |  | Consider pre-processing of Eigenvector using Enhanced Type 2 codebook |  | Eigen vectors, Sub-band reporting (e.g., [13] sub-bands for 10 MHz CBW, 15kHz SCS). |  | Eigen vectors | Eigenvector |
| Training procedure | FFS (e.g Initialization method, training duration, training completion criteria, collaboration type, encoder assumption, etc) Note that training procedure does not need to be fully aligned. |  |  |  |  | Convolutional: Feedback bits per transmission e.g., 10 x 4 = 40 Transformer: Optimizer, e.g., Adam | collaboration training type need to be specified | Collaboration type: Type-3 Network first training | Training completion criteria is probably one of the most important parameters assuming it will include boundaries (minimum and maximum) for the test decoder performance required. Collaboration type will determine interactions required between different stakeholders and/or different AI/ML algorithms blocks |  |  |
| Loss function | SGCS, NMSE, etc. |  | SGCS |  |  | NMSE | SGCS, NMSE, etc. |  |  |  |  |
| Training datasets | Channel model, number of Tx/Rx ports Other parameters FFS (e.g. rank) Dataset containing only channel information, which is merged by data from companies. |  | Encoder input dataset should cover all the contributing companies’ encoder input data |  |  |  | Number of layers/rank? SNR, Genie/ real channel estimates (impairments)? Data format of training (depends on Collaboration training type)  Size of training data set Specify channel model parameters or training data samples stored in a repository? Different Training Sets (configurations/ scenarios)? Multiple vendor training sets | Channel model for training: UMa Note that in the performance test TDL or CDL (if available) model to be used. Number of Tx/Rx ports: 4 RX, 16 or 32 TX, Note that other options should not be precluded but better to agree on a single scenario as a starting point. Rank: 1. Channel estimates: Channel eigenvectors derived from [ideal, non-ideal] channel estimates, magnitude normalized to unit length. Dataset size: Sufficient number of samples to achieve minimum performance and prevent underfitting are needed. |  |  |  |
| Hyperparameters | Learning rate, batch size, regularization techniques and strength, optimization algorithm, etc. Note that training procedure does not need to be fully aligned. |  |  | Learning rate = 0.001, batch size = 128 optimization algorithm = Adam, |  |  | Learning rate, batch size, regularization techniques and strength, optimization algorithm, etc. |  |  |  |  |
| Cross-validation details | Dataset splits for training/testing/validation |  |  | Dataset for training: 199,500 Dataset for Testing: 10000 Dataset for validation: 10500 |  |  | Dataset splits for training/validation/testing This testing doesn’t refer to DUT testing | 80%/20%, where training data is also used for validation. |  |  |  |

* Recommended WF
  + To be discussed: which parameters are needed and what should the values be

### Sub-topic 4-4

*Simulation parameters for Option 3*

Several companies brought proposals on simulation parameters to be used for encoder/decoder derivation to further progress the feasibility of option 3. These should be discussed and a set of parameters should be agreed.

**Issue 4-4: Simulation parameters for Option 3**

* Proposals are shown in the table below

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | CATT | Nokia | Intel | Vivo | QC |
| Duplex, Waveform | FDD, OFDM | FDD OFDM | FDD | FDD OFDM | FDD OFDM |
| Carrier frequency | 2GHz | 2GHz |  | 2GHz |  |
| Bandwidth | 10MHz | 40MHz | 10 MHz | 20MHz | 20MHz |
| Subcarrier spacing | 15kHz for 2GHz | 30kHz | 15 kHz | 15kHz | 15kHz |
| Nt | 32: (8,8,2,1,1,2,8), (dH,dV) = (0.5, 0.8)λ | 32: (8,4,2,1,1,4,4), (dH,dV) = (0.5, 0.8)λ and/or 16: (8,4,2,1,1,2,4), (dH,dV) = (0.5, 0.8)λ | CDL channel models - 32 ports: (8,8,2,1,1,2,8), (dH,dV) = (0.5, 0.8)λ - 16 ports: (8,4,2,1,1,2,4), (dH,dV) = (0.5, 0.8)λ TDL channel models - Low correlation | 32: (8,8,2,1,1,2,8), (dH,dV) = (0.5, 0.8)λ | 32: (8,8,2,1,1,2,8), (dH,dV) = (0.5, 0.8)λ |
| Nr | 4: (1,2,2,1,1,1,2), (dH,dV) = (0.5, 0.5)λ | 4: (1,2,2,1,1,1,2), (dH,dV) = (0.5, 0.5)λ | CDL channel models - 4RX UE: (1,2,2,1,1,1,2), (dH,dV) = (0.5, 0.5)λ for (rank 1-4) TDL channel models - 4RX UE, Low correlation | 4: (1,2,2,1,1,1,2), (dH,dV) = (0.5, 0.5)λ | 4: (1,2,2,1,1,1,2), (dH,dV) = (0.5, 0.5)λ |
| Channel model (Delay spread) | CDL-C (300 ns) as baseline, CDL-A (30 ns) as optional. | TDLA30-5 as starting point, CDL-C (3km UE speed, 30n delay spread) can be considered | 1) CDL-C 300ns DS, 100Hz (TR 38.901), 2) TDLC300-100 (TS 38.101-4) | CDL-C | CDL-C or TDL-A |
| UE speed | 3kmhr |  |  | 3kmhr | 10Hz doppler |
| Channel estimation | Realistic channel estimation algorithms (e.g., LS or MMSE) as a baseline. Ideal DL channel estimation is optionally taken into the baseline of evaluation methodology for the purpose of calibration and/or comparing intermediate results (e.g., accuracy of AI/ML output CSI, etc.). Up to companies to report whether/how ideal channel is used in the dataset construction and performance evaluation/inference. | Realistic channel estimation algorithms (e.g., LS or MMSE) as a baseline. | Realistic channel estimation algorithms  MMSE-IRC as the baseline receiver | Realistic channel estimation algorithms (e.g., LS or MMSE) as a baseline. Ideal DL channel estimation is optionally taken into the baseline of evaluation methodology for the purpose of calibration and/or comparing intermediate results (e.g., accuracy of AI/ML output CSI, etc.). Up to companies to report whether/how ideal channel is used in the dataset construction and performance evaluation/inference. | Realistic channel estimation algorithms (e.g., LS or MMSE) or ideal DL channel estimation |
| Rank per UE | Rank 1. | Rank 1 | Fixed rank 1 as the starting point 4 PRB subband size | Rank 1-4. Companies are encouraged to report the Rank number, and whether/how rank adaptation is applied | Rank 1 |
| Evaluation metric(s) | SGCS |  | SGCS metric (as described in TR 38.843 6.2.1) or NMSE |  |  |
| Delay spread |  |  |  | 30ns | 30ns |
| Latent message size |  |  |  |  | Use power of 2, choose from 32, 64, 128, 256 and 512 bits. |

* Recommended WF
  + Parameters to be discussed and agreed

### Sub-topic 4-5

*Option 4 for 2-sided model*

Several companies brought proposal on how to further study/check the feasibility of option 4.

**Issue 4-6: Option 4 for 2-sided model**

* Proposals
  + Option 1: Qualcomm (R4-2407334)
* **Option 4a-1 standardized dataset**
  + **Step 1: RAN4 agrees a pair of encoder and decoder with full details (same as fully specified decoder discussion) and an encoder input data generation procedure.**
  + **Step 2: RAN4 uses this encoder/decoder pair and the generation procedure to generate a set of decoder input and output data and captures this dataset in the specification.**
  + **Step 3: RAN4 specifies a test decoder verification procedure based on the specified dataset.**
* **Option 4a-2 standardized aggregated dataset**
  + **Step 1: RAN4 achieves some agreements (e.g., part of but not all the parameters in the test decoder parameter table in the previous meeting WF[1]) for the test decoder.**
  + **Step 2: Interested companies can design their own encoder/decoder pairs based on the agreements to contribute the (decoder input, decoder output) dataset to RAN4**
  + **Step 3: RAN4 aggregates the datasets from all the contributing companies, and capture the aggregated dataset in the specification**
  + **Step 4: RAN4 specifies a test decoder verification procedure based on the specified dataset.**
* **Option 4b reference encoder/decoder pair**
  + **Step 1: RAN4 agrees a pair of encoder and decoder with full details (same as fully specified decoder discussion) and an encoder input data generation procedure**
  + **Step 2: RAN4 capture the encoder/decoder as a reference encoder/decoder pair and the encoder input data generation procedure in the specification.**
  + **Step 3: RAN4 specifies a test decoder verification procedure based on the reference encoder.**
* Option 2: Xiaomi (R4-2407847): ***Standardized data / dataset format + Dataset exchange between NW-side and UE-side***
* Option 3: vivo(R4-2408294)
  + **Option 4a: Reference encoder + test decoder model structure, and channel generation method are in the spec;**
  + **Option 4b: Test decoder model structure + reference encoder model structure, and dataset (PMI and corresponding channel) are in the spec;**
  + Option 4 – Huawei (R4-2409003))
* Option 4a: Model structure is not specified in RAN4. Training dataset is specified, where each training sample consists of both the raw channel matric/precoding matrix and the bit stream forwarded to the test decoder.
* Option 4b: Model structure is specified in RAN4. Training dataset is not specified for verifying the encoder at DUT. The test decoder developed by TE vendor needs verification.
  + FFS: How to determine the test metric for test decoder developed by each TE vendor.
  + Option 5: Nokia (R4-2408659)
* **Option 2 (Dataset based):**
  + **Option 2a: Freeze complete training data while leaving model architecture for implementation.**
  + **Option 2b: Freeze the important characteristics of training data, e.g., number of bits of latent message while leaving actual data samples and model architecture for implementation.**
* **Option 3: Freeze the important characteristics of training data, e.g., number of bits of latent message, and a backbone of model architecture while leaving actual data samples and architectural details for implementation.**
  + Option 6: Ericsson (R4-2408492)
    - the latent space needs to be standardized
    - standardize a reference encoder in order to capture the latent space
  + Option 7: others
* Recommended WF
  + To be discussed

Likely multiple options need to be chosen/combined RAN4 should agree on a minimum set such that companies can continue the study

### Sub-topic 4-6

*TE verification/validation for Option 4*

Option will leave the actual test decoder implementation to the TE vendors, there might be a need to verify that the decoder is correctly implemented and has the right level of performance.

**Issue 4-6: TE decoder verification/validation**

* Proposals
  + Option 1: RAN4 will have to come up with a TE verification/validation procedure
    - details are FFS
  + Option 2: This should be left to RAN5
  + Option 3: TE verification/validation is not needed
  + Option 4: other options
* Recommended WF

To be discussed

# References

[1] R4-2408944, “Topic summary for [111][133] NR\_AIML\_air”, Moderator(Qualcomm Incorporated), RAN4#111