**3GPP TSG-RAN WG1 Meeting #118 R1-** **24xxxxx**

**Maastricht, Netherland, August 19th – August 23th, 2024**

**Agenda item:** 9.1.3.2

**Source:** Moderator (Qualcomm)

**Title:** Summary of Evaluation Results for Additional study on AI/ML for NR air interface: CSI compression

**Document for:** Discussion and Decision

# Introduction

This documents contains a summary of evaluation results to be captured into TR.

## Contact Information

Please provide / update your contact information.

|  |  |  |
| --- | --- | --- |
| **Company** | **Name** | **E-mail** |
| Moderator (Qualcomm) | Taesang Yoo | taesangy@qti.qualcomm.com |
| Lenovo | Vahid Pourahmadi | vpourahmadi@lenovo.comwangjf20@lenovo.com |
| New H3C | Lei Zhou | Zhou.leih@h3c.com |
| Samsung | Ameha Tsegaye Abebe | amehat.abebe@samsung.com |
| Indian Institute of Technology Madras | Anil Kumar Yerrapragada | anilkumar@5gtbiitm.in |
| Fujitsu | WANG Guotong (David) | wangguotong@fujitsu.com |
| vivo | Peng SUN | sunpeng@vivo.com |
| SK telecom | Yunesung Kim | yunesung.kim@sktelecom.com |
| Mavenir | Ali Fatih Demir | ali.demir@mavenir.com |
| Panasonic | Tetsuya YamamotoHidetoshi Suzuki | yamamoto.tetsuya001@jp.panasonic.comsuzuki.hidetoshi@ jp.panasonic.com |
| OPPO | Wendong Liu | liuwendong1@oppo.com |
| NTT DOCOMO | Xin WangHaruhi Echigo | wangx@docomolabs-beijing.com.cnharuhi.echigo.fw@nttdocomo.com |
| ETRI | Anseok Lee | alee@etri.re.kr |
| Ericsson | Jingya LiSiva MuruganathanJianwei Zhang | Jingya.li@ericsson.comSiva.muruganathan@ericsson.comJianwei.zhang@ericsson.com |
| ZTE | Lun LiXingguang Wei | li.lun1@zte.com.cnwei.xingguang@zte.com.cn |
| IIT Kanpur | Shyam Vijay GadhaiAbhishek Kumar Singh | svgadhai@iitk.ac.inabhishekks@iitk.ac.in |
| Spreadtrum | Mimi Chen | Mimi.chen@unisoc.com |
| LG | Jaehoon Chung | jhoon.chung@lge.com |
| CATT | Qianrui Li | liqianrui@catt.cn |
| Intel | Victor Sergeev | Victor.sergeev@intel.com |
| Xiaomi | LiuzhengxuanLiumin | liuzhengxuan@xiaomi.comliumin10@xiaomi.com |
| AT&T | Isfar TariqSalam Akoum | Isfar.tariq@att.comSalam.akoum@att.com |
| CMCC | Yuhua CaoYi ZhengDan Song | caoyuhua@chinamobile.comzhengyi@chinamobile.comsongdan@chinamobile.com |
| NVIDIA | Xingqin Lin | xingqinl@nvidia.com |
| MediaTek | Pedram Kheirkhah Sangdeh | Pedram.kheirkhah@mediatek.com |
| Futurewei | Baoling S Sheen | bsheen@futurewei.com |
| TCL | Yunsheng Kuang | yunsheng.kuang@tcl.com |
| Huawei, HiSilicon | Yuan Li | liyuan3@huawei.com |
| CEWiT(TSDSI) | Shiv ShankarDhivagar Baskaran | shivshankar@cewit.org.indhivagar.b@cewit.org.in |
| NEC | Zhen HePravjyot Deogun | he\_zhen@nec.cnpravjyot.deogun@EMEA.NEC.COM |
| Qualcomm | Chenxi Hao | chenxih@qti.qualcomm.com |
| Sony | Hiroki MatsudaSam Atungsiri | hiroki.matsuda@sony.comsam.atungsiri@sony.com |
| Apple | Huaning Niu | Huaning\_niu@apple.com |
| Tejas Networks | Pavan Kalyan | pavankalyand@tejasnetworks.com |

# Texts and plots to be captured into TR

The description of the above boxcharts is as follows:

* The line inside of each box is the median
* The top and bottom box edges represent the 0.75 quantile (upper quartile) and 0.25 quantile (lower quartile), respectively.
* Interquartile range (IQR) is the distance between the top and bottom box edges, which is used to determine the outliers (values more than 1.5 IQR away from the box edge).
* Outliers are represented by the ‘o’ marks.
* The whiskers represents the minimum and maximum values excluding outliers
* 1, 2, 3 on the x-axis represents either the three payload ranges X, Y, Z or A, B, C, or the low (RU<=39%), medium (RU 40-69%), and high (RU>=70%) Resource Utilization regions.

Observation on SGCS performance Case 1

For the evaluation of temporal domain aspects **Case 1** of AI/ML based CSI compression compared to the non-AI/ML *benchmark in terms of SGCS*,

* For Layer 1,
	+ 1 source [OPPO] observes performance gain of 4.9% at CSI payload X (small payload);
	+ 1 source [CMCC] observes performance gain of 29.94% at CSI payload Y (medium payload).
	+ Performance gain at CSI payload Z (large payload) is TBD
	+ 1 source [MediaTek] observes performance gain of 3.0% at CSI payload Z (large payload)

For the evaluation of temporal domain aspects **Case 1** of AI/ML based CSI compression compared to the CSI compression Case 0 *in terms of SGCS*,

* For Layer 1,
	+ 2 sources [Futurewei, OPPO] observe performance gain of -3.2% to 12.1% at CSI payload X (small payload)
	+ 2 sources [Futurewei, ~~OPPO~~, CMCC] observe performance gain of 6.95-12% at CSI payload Y (medium payload)
	+ 1 source [MediaTek] observes performance gain of 0.7% at CSI payload Z (large payload)
	+ Performance gain at CSI payload Z (large payload) is TBD

Observation on FTP traffic performance Case 1

For the evaluation of temporal domain aspects **Case 1** of AI/ML based CSI compression compared to the non-AI/ML *benchmark in terms of mean UPT under FTP traffic*, till RAN1 #118,

* For Max Rank 1, performance gains are TBD
* For Max Rank 2,
	+ For RU <=39%, performance gains are TBD
	+ For RU of 40-69%, performance gains are TBD.
	+ For RU >= 70%, 1 source [Futurewei] observes performance gains of 26%
		- 1 source [Futurewei] observes performance gains of 26% at CSI feedback overhead A (small overhead)
		- TBD performance gains at CSI feedback overhead B (medium overhead)
		- TBD performance gains at CSI feedback overhead C (large overhead)
* For Max Rank 4, performance gains are TBD.

For the evaluation of temporal domain aspects **Case 1** of AI/ML based CSI compression compared to the Case 0 *benchmark in terms of mean UPT under FTP traffic*, till RAN1 #118,

* For Max Rank 1, performance gains are TBD
* For Max Rank 2,
	+ For RU <=39%, performance gains are TBD
	+ For RU of 40-69%, performance gains are TBD.
	+ For RU >= 70%, 1 source [Futurewei] observes performance gains of 2%
		- 1 source [Futurewei] observes performance gains of 2% at CSI feedback overhead A (small overhead)
		- TBD performance gains at CSI feedback overhead B (medium overhead)
		- TBD performance gains at CSI feedback overhead C (large overhead)
* For Max Rank 4, performance gains are TBD.

For the evaluation of temporal domain aspects **Case 1** of AI/ML based CSI compression compared to the non-AI/ML *benchmark in terms of 5% UPT under FTP traffic*, till RAN1 #118,

* For Max Rank 1, performance gains are TBD
* For Max Rank 2,
	+ For RU <=39%, performance gains are TBD
	+ For RU of 40-69%, performance gains are TBD.
	+ For RU >= 70%, 1 source [Futurewei] observes performance gains of 73%
		- 1 source [Futurewei] observes performance gains of 73% at CSI feedback overhead A (small overhead)
		- TBD performance gains at CSI feedback overhead B (medium overhead)
		- TBD performance gains at CSI feedback overhead C (large overhead)
* For Max Rank 4, performance gains are TBD.

For the evaluation of temporal domain aspects **Case 1** of AI/ML based CSI compression compared to the Case 0 *benchmark in terms of 5% UPT under FTP traffic*, till RAN1 #118,

* For Max Rank 1, performance gains are TBD
* For Max Rank 2,
	+ For RU <=39%, performance gains are TBD
	+ For RU of 40-69%, performance gains are TBD.
	+ For RU >= 70%, 1 source [Futurewei] observes performance gains of 12%
		- 1 source [Futurewei] observes performance gains of 12% at CSI feedback overhead A (small overhead)
		- TBD performance gains at CSI feedback overhead B (medium overhead)
		- TBD performance gains at CSI feedback overhead C (large overhead)
* For Max Rank 4, performance gains are TBD.

Observation on Full buffer Case 1

Observation on CSI feedback reduction Case 1

For the evaluation of temporal domain aspects **Case 1** of AI/ML based CSI compression, compared to the non-AI/ML benchmark, in terms of CSI feedback reduction, till RAN1 #118,

* For Max rank = 1, CSI feedback reduction is TBD.
* For Max rank = 2,
	+ For CSI feedback overhead A (small overhead), 1 source [Futurewei] observes CSI feedback reduction of 92%
* For Max rank = 4, CSI feedback reduction is TBD.

For the evaluation of temporal domain aspects **Case 1** of AI/ML based CSI compression, compared to the Case 0 benchmark, in terms of CSI feedback reduction, till RAN1 #118,

* CSI feedback reduction is TBD.

The above results are based on the following assumptions besides the assumptions of the agreed EVM table:

* Precoding matrix is used as the model input.
* Training data samples are not quantized, i.e., Float32 is used/represented.
* 1-on-1 joint training is assumed.
* The performance metric is SGCS for Layer 1 of Max rank 1 or Layer 1/2 of Max rank 2.
* CSI payload X is ≤ 80/α bits; CSI payload Y is (100 - 140 )/α bits; CSI payload Z is ≥ 230/α bits; where X, Y, Z are applicable per layer, where alpha = 1 for rank = 1/2 and alpha = 2 for rank = 3/4.
* Benchmark is Rel-16 Type II codebook.

Observation on SGCS performance Case 2

For the evaluation of temporal domain aspects Case 2 of AI/ML based CSI compression compared to the non-AI/ML *benchmark in terms of SGCS*,

For Layer 1,

* 13 sources [Fujitsu, ZTE, Apple, QC, Samsung, vivo, OPPO, Xiaomi, Spreadtrum, Huawei, ETRI, Nokia, Futurewei] observe performance gain of 9.12-27.8% at CSI payload X (small payload), for which the median SGCS gain is 14.8%;
* 10 sources [ZTE, QC, vivo, CATT, CMCC, Xiaomi, Spreadtrum, Huawei, ETRI, Nokia] observe performance gain of 4.34-27.9% at CSI payload Y (medium payload), for which the median SGCS gain is 11%;
* 8 sources [ZTE, QC, vivo, Huawei, CATT, Xiaomi, Spreadtrum, Nokia] observes performance gain of 3.4-17.2% at CSI payload Z (large payload), for which the median SGCS gain is 6.26%.
* The following boxchart shows the median, 0.75 quantile, 0.25 quantile, outliers, and min/max values excluding outliers, for (X,Y,Z) CSI payload bins.



For Layer 2,

* 8 sources [ZTE, QC, Samsung, vivo, Huawei, Xiaomi, Nokia, Futurewei] observe performance gain between 14.2-37.5% at CSI payload X (small payload) , for which the median SGCS gain is 19.67%.
* 6 sources [ZTE, QC, vivo, Huawei, Nokia, Xiaomi] observe performance gain of 3.22-30% at CSI payload Y (medium payload), for which the median SGCS gain is 17%.
* 6 sources [ZTE, QC, vivo, Huawei, Nokia, Xiaomi] observe performance gain of 5.17-18% at CSI payload Z (large payload) , for which the median SGCS gain is 13.07%.
* The following boxchart shows the median, 0.75 quantile, 0.25 quantile, outliers, and min/max values excluding outliers, for (X,Y,Z) CSI payload bins.



For Layer 3,

* 2 sources [QC, Samsung] observe performance gain between 29.95-146.8% at CSI payload X (small payload)
* 1 source [QC] observes performance gain of 44.9% at CSI payload Y (medium payload)
* 1 source [QC] observes performance gain of 23.7% at CSI payload Z (large payload)

For Layer 4,

* 2 sources [QC, Samsung] observe performance gain between 38.55-280.6% at CSI payload X (small payload)
* 1 source [QC] observes performance gain of 59% at CSI payload Y (medium payload)
* 1 source [QC] observes performance gain of 33.5% at CSI payload Z (large payload)

For the evaluation of temporal domain aspects **Case 2** of AI/ML based CSI compression compared to the CSI compression Case 0 *in terms of SGCS*,

For Layer 1,

* 14 sources [Fujitsu, ZTE, Apple, QC, Samsung, vivo, OPPO, Xiaomi, Spreadtrum, Huawei, ETRI, Nokia, Futurewei, InterDigital] observe performance gain of 0.62-30% at CSI payload X (small payload) , for which the median SGCS gain is 8.6%.
* 11 sources [ZTE, QC, vivo, CATT, CMCC, Xiaomi, Spreadtrum, Huawei, ETRI, Nokia, IIT Kanpur] observe performance gain of 1.49-36% at CSI payload Y (medium payload) , for which the median SGCS gain is 6.02%.
* 9 sources [ZTE, QC, vivo, Huawei, CATT, Xiaomi, Spreadtrum, Nokia, IIT Kanpur] observe performance gain of 0.5-6.78% at CSI payload Z (large payload) , for which the median SGCS gain is 4.3%.
* The following boxchart shows the median, 0.75 quantile, 0.25 quantile, outliers, and min/max values excluding outliers, for (X, Y, Z) CSI payload bins.



For Layer 2,

* 9 sources [ZTE, QC, Samsung, vivo, Huawei, Xiaomi, Nokia, Futurewei, InterDigital] observe performance gain of 0.61-20% at CSI payload X (small payload), for which the median SGCS gain is 9.73%.
* 6 sources [ZTE, QC, vivo, Huawei, Nokia, Xiaomi] observe performance gain of 1.06-16.49% at CSI payload Y (medium payload) , for which the median SGCS gain is 10%.
* 6 sources [ZTE, QC, vivo, Huawei, Nokia, Xiaomi] observe performance gain of 2.98-13% at CSI payload Z (large payload) , for which the median SGCS gain is 7.95%.
* The following boxchart shows the median, 0.75 quantile, 0.25 quantile, outliers, and min/max values excluding outliers, for (X, Y, Z) CSI payload bins.



For Layer 3,

* 2 source [QC, Samsung] observe performance gain between 8.7-29.96% at CSI payload X (small payload)
* 1 source [QC] observes performance gain of 8.4% at CSI payload Y (medium payload)
* 1 source [QC] observes performance gain of 7.9% at CSI payload Z (large payload)

For Layer 4,

* 2 source [QC, Samsung] observes performance gain between 8.33-10.9% at CSI payload X (small payload)
* 1 source [QC] observes performance gain of 8.7% at CSI payload Y (medium payload)
* 1 source [QC] observes performance gain of 7.6% at CSI payload Z (large payload)

The above results are based on the following assumptions besides the assumptions of the agreed EVM table:

* Precoding matrix (SVD output or in angle-delay domain) is used as the model input.
* Training data samples are not quantized, i.e., Float32 is used/represented.
* 1-on-1 joint training is assumed.
* The performance metric is SGCS for Layer 1 of Max rank 1, Layer 1/2 of Max rank 2, Layer 1/2/3/4 of Max Rank 4.
* CSI payload X is ≤ 80/α bits; CSI payload Y is (100 - 140 )/α bits; CSI payload Z is ≥ 230/α bits; where X, Y, Z are applicable per layer, where alpha = 1 for rank = 1/2 and alpha = 2 for rank = 3/4.

Observation on FTP traffic performance Case 2

For the evaluation of temporal domain aspects **Case 2** of AI/ML based CSI compression compared to the non-AI/ML *benchmark in terms of mean UPT* *under FTP* traffic, till RAN1 #118,

For Max Rank 1,

* For RU <= 39%, 2 sources [Huawei, QC] observes performance gain of 0-3.4%:
	+ 2 sources [Huawei, QC] observes performance gain of 0-3.4% at CSI feedback overhead A (small overhead)
	+ 2 sources [Huawei, QC] observes performance gain of 0-2.4% at CSI feedback overhead B (medium overhead)
	+ 2 source [Huawei, QC] observes performance gain if 0-2% at CSI feedback overhead C (large overhead)
* For RU 40-69%, 3 sources [Huawei, Spreadtrum, QC] observed performance gain of 2-6%
	+ 3 sources [Huawei, Spreadtrum, QC] observe performance gain of 3-6% at CSI feedback overhead A (small overhead)
	+ 3 sources [Huawei, Spreadtrum, QC] observe performance gain of 2-5% at CSI feedback overhead B (medium overhead)
	+ 3 sources [Huawei, Spreadtrum, QC] observe performance gain of 2-6% at CSI feedback overhead C (large overhead)
* For RU > 70%, 4 sources [Huawei, Spreadtrum, QC, Oppo] observes performance gain of 3-15%
	+ 4 sources [Huawei, Spreadtrum, QC, Oppo] observed performance gain of 3-15% at CSI feedback overhead A (small overhead)
	+ 4 sources [Huawei, Spreadtrum, QC, Oppo] observes performance gain of 4-8% at CSI feedback overhead B (medium overhead)
	+ 4 sources [Huawei, Spreadtrum, QC, Oppo] observes performance gain of 4-8% at CSI feedback overhead C (large overhead)

For Max Rank 2,

* For RU <= 39%, 5 sources [Huawei, Interdigital, QC, Nokia, ZTE] observes performance gain of 0-12%:
	+ 5 sources [Huawei, Interdigital, QC, Nokia, ZTE] observe performance gain of 1-12% at CSI feedback overhead A (small overhead), for which the performance gain is 7.3%.
	+ 4 sources [Huawei, QC, Nokia, ZTE] observes performance gain of 0-3.3% at CSI feedback overhead B (medium overhead)
	+ 4 source [Huawei, QC, Nokia, ZTE] observes performance gain if 0-5.6% at CSI feedback overhead C (large overhead)
* For RU 40-69%, 5 sources [Huawei, Interdigital, QC, Nokia, ZTE] observed performance gain of 1-17%
	+ 5 sources [Huawei, Interdigital, QC, Nokia, ZTE] observe performance gain of 4-17% at CSI feedback overhead A (small overhead), for which the performance gain is 12.8%.
	+ 4 sources [Huawei, QC, Nokia, ZTE] observes performance gain of 2-11% at CSI feedback overhead B (medium overhead)
	+ 4 sources [Huawei, QC, Nokia, ZTE] observes performance gain of 1-9% at CSI feedback overhead C (large overhead)
* For RU > 70%, 6 sources [Huawei, Interdigital, Futurewei, QC, Nokia, ZTE] observe performance gain of 3-29%
	+ 6 sources [Huawei, Interdigital, Futurewei, QC, Nokia, ZTE] sources observed performance gain of 6-29% at CSI feedback overhead A (small overhead), for which the performance gain is 16%.
	+ 4 sources [Huawei, QC, Nokia, ZTE] observes performance gain of 3-17% at CSI feedback overhead B (medium overhead)
	+ 4 source [Huawei, QC, Nokia, ZTE] observes performance gain of 3-17% at CSI feedback overhead C (large overhead)

The following boxchart shows the median, 0.75 quantile, 0.25 quantile outliers and min/max values excluding outliers, for “A” CSI payload bins, for various resource utilizations. Multiple submissions by a few companies are why we have more points than the number of companies above.



For Max Rank 4,

* For RU <= 39%, 1 source [QC] observes a performance gain of 0-6%
	+ 1 source [QC] observes a performance gain of 5.5% at CSI feedback overhead A (small overhead).
	+ 1 source [QC] observes a performance gain of 5.3% at CSI feedback overhead B (medium overhead).
	+ 1 source [QC] observes a performance gain of 0.1% at CSI feedback overhead C (large overhead).
* For RU 40-69%, 1 source [QC] observes a performance gain of 8-18%
	+ 1 source [QC] observes a performance gain of 17.5% at CSI feedback overhead A (small overhead).
	+ 1 source [QC] observes a performance gain of 16.4% at CSI feedback overhead B (medium overhead).
	+ 1 source [QC] observes a performance gain of 8.8% at CSI feedback overhead C (large overhead).
* For RU > 70%, 1 source [QC] observes a performance gain of 12-24%
	+ 1 source [QC] observes a performance gain of 23.5% at CSI feedback overhead A (small overhead).
	+ 1 source [QC] observes a performance gain of 22.7% at CSI feedback overhead B (medium overhead).
	+ 1 source [QC] observes a performance gain of 12.9% at CSI feedback overhead C (large overhead).

For the evaluation of temporal domain aspects **Case 2** of AI/ML based CSI compression compared to the CSI compression Case 0 *~~benchmark~~ in terms of mean UPT* *under FTP* traffic, till RAN1 #118,

For Max Rank 1,

* For RU <= 39%, 1 source [Huawei] observes performance gain of 0.8-1.6%:
	+ 1 source [Huawei] observes performance gain of 1.6% at CSI feedback overhead A (small overhead)
	+ 1 source [Huawei] observes performance gain of 1.4% at CSI feedback overhead B (medium overhead)
	+ 1 source [Huawei] observes performance gain if 0.8% at CSI feedback overhead C (large overhead)
* For RU 40-69%, 3 sources [Huawei, Spreadtrum, QC] observed performance gain of 0-3%
	+ 3 sources [Huawei, Spreadtrum, QC] observes performance gain of 0-3% at CSI feedback overhead A (small overhead)
	+ 3 sources [Huawei, Spreadtrum, QC] observes performance gain of 0-2% at CSI feedback overhead B (medium overhead)
	+ 3 sources [Huawei, Spreadtrum, QC] observes performance gain of 0-1.4% at CSI feedback overhead C (large overhead)
* For RU > 70%, 4 sources [Huawei, Spreadtrum, QC, Oppo] observes performance gain of 0-6%
	+ 4 sources [Huawei, Spreadtrum, QC, Oppo] observed performance gain of 0-6% at CSI feedback overhead A (small overhead)
	+ 3 sources [Huawei, Spreadtrum, QC] observes performance gain of 2-3.3% at CSI feedback overhead B (medium overhead)
	+ 3 sources [Huawei, Spreadtrum, QC] observes performance gain of 0.9-1.4% at CSI feedback overhead C (large overhead)

For Max Rank 2,

* For RU <= 39%, 5 sources [Huawei, Interdigital, QC, Nokia, ZTE] observe performance gain of -1% to 3.3%:
	+ 4 sources[Huawei, Interdigital, Nokia, ZTE] observe performance gain of -1% to 3.3% at CSI feedback overhead A (small overhead), for which the performance gain is 1.09%.
	+ 3 source [Huawei, Nokia, ZTE] observes performance gain of 0.17%-2% at CSI feedback overhead B (medium overhead)
	+ 3 sources [Huawei, Nokia, ZTE] observes performance gain if -0.05% to 2% at CSI feedback overhead C (large overhead)
* For RU 40-69%, 6 sources [Huawei, Interdigital, QC, Nokia, Futurewei, ZTE] observed performance gain of -2% to 7%
	+ 6 sources [Huawei, Interdigital, QC, Nokia, Futurewei, ZTE] observe performance gain of -2% to 7% at CSI feedback overhead A (small overhead), for which the performance gain is 3.375%.
	+ 4 source [Huawei, QC, Nokia, ZTE] observes performance gain of 0-4.2% at CSI feedback overhead B (medium overhead)
	+ 4 source [Huawei, QC, Nokia, ZTE] observes performance gain of 0-3% at CSI feedback overhead C (large overhead)
* For RU > 70%, 6 sources [Huawei, Futurewei, Interdigital, QC, Nokia, ZTE] observes performance gain of -5% to 14%
	+ 6 sources Huawei, Futurewei, Interdigital, QC, Nokia, ZTE] observed performance gain of -5% to 14% at CSI feedback overhead A (small overhead), for which the performance gain is 3.96%.
	+ 4 sources [Huawei, QC, Nokia, ZTE] observes performance gain of 1-12% at CSI feedback overhead B (medium overhead)
	+ 4 sources [Huawei, QC, Nokia, ZTE] observes performance gain of 0-10% at CSI feedback overhead C (large overhead)

The following boxchart shows the median, 0.75 quantile, 0.25 quantile, outliers, and min/max values excluding outliers, for “A” CSI payload bins for various resource utilizations. Multiple submissions by a few companies are why we have more points than the number of companies above.



For Max Rank 4,

* For RU <= 39%, 1 source [QC] observes performance gain of -0.2% to 0.9%:
	+ 1 source [QC] observes a performance gain of -0.2% at CSI feedback overhead A (small overhead).
	+ 1 source [QC] observes a performance gain of 0.5% at CSI feedback overhead B (medium overhead).
	+ 1 source [QC] observes a performance gain of 0.9% at CSI feedback overhead C (large overhead).
* For RU 40-69%, 1 source [QC] observes a performance gain of 4.2-6.5%
	+ 1 source [QC] observes a performance gain of 6.5% at CSI feedback overhead A (small overhead).
	+ 1 source [QC] observes a performance gain of 6.0% at CSI feedback overhead B (medium overhead).
	+ 1 source [QC] observes a performance gain of 4.2% at CSI feedback overhead C (large overhead).
* For RU > 70%, 1 source [QC] observes a performance gain of 7.5-11.1%
	+ 1 source [QC] observes a performance gain of 11.1% at CSI feedback overhead A (small overhead).
	+ 1 source [QC] observes a performance gain of 10.9% at CSI feedback overhead B (medium overhead).
	+ 1 source [QC] observes a performance gain of 7.5% at CSI feedback overhead C (large overhead)..

For the evaluation of temporal domain aspects **Case 2** of AI/ML based CSI compression compared to the non-AI/ML *benchmark in terms of 5% UPT under FTP*, till RAN1 #118,

For Max rank 1:

* For RU <= 39%, 2 sources [Huawei, QC] observes performance gain of 0-8%:
	+ 2 sources [Huawei, QC] observe the performance gain of 1.5-8% at CSI feedback overhead A (small overhead)
	+ 2 sources [Huawei, QC] observes the performance gain of 0.3%-4% at CSI feedback overhead B (medium overhead)
	+ 2 sources [Huawei, QC] observes the performance gain of 0.7-4% at CSI feedback overhead C (large overhead)
* For RU between 40-69%, 3 sources [Huawei, Spreadtrum, QC] observes performance gain of 4-12%:
	+ 3 sources [Huawei, Spreadtrum, QC] observes the performance gain of 10-12% at CSI feedback overhead A (small overhead)
	+ 3 sources [Huawei, Spreadtrum, QC] observes the performance gain of ~8% at CSI feedback overhead B (medium overhead)
	+ 3 source [Huawei, Spreadtrum, QC] observes the performance gain of 4-7% at CSI feedback overhead C (large overhead)
* For RU > 70%, 3 sources [Huawei, Spreadtrum, QC] observes performance gain of 10-37%:
	+ 3 sources [Huawei, Spreadtrum, QC] observes the performance gain of 10-37% at CSI feedback overhead A (small overhead)
	+ 3 sources [Huawei, Spreadtrum, QC] observes the performance gain of 12-27% at CSI feedback overhead B (medium overhead)
	+ 3 sources [Huawei, Spreadtrum, QC] observes the performance gain of 10-30% at CSI feedback overhead C (large overhead)

For Max Rank 2:

* For RU <= 39%, 5 sources [Huawei, Interdigital, QC, Nokia, ZTE] observe performance gain of 1-45%:
	+ 5 sources [Huawei, Interdigital, QC, Nokia, ZTE] observe the performance gain of 6-45% at CSI feedback overhead A (small overhead) , for which the performance gain is 13.7%.
	+ 4 sources [Huawei, QC, Nokia, ZTE] observes the performance gain of 1.5-14% at CSI feedback overhead B (medium overhead)
	+ 4 sources [Huawei, QC, Nokia, ZTE] observes the performance gain of 1-8% at CSI feedback overhead C (large overhead)
* For RU between 40-69%, 5 sources [Huawei, Interdigital, QC, Nokia, ZTE] observe performance gain of 3-41%:
	+ 5 sources [Huawei, Interdigital, QC, Nokia, ZTE] observe the performance gain of 9-41% at CSI feedback overhead A (small overhead) , for which the performance gain is 26%.
	+ 4 sources [Huawei, QC, Nokia, ZTE] observes the performance gain of 3-33% at CSI feedback overhead B (medium overhead)
	+ 4 sources [Huawei, QC, Nokia, ZTE] observes the performance gain of 3-21% at CSI feedback overhead C (large overhead)
* For RU > 70%, 6 sources [Huawei, Futurewei, Interdigital, QC, Nokia, ZTE] observe performance gain of 6-73%:
	+ 6 sources [Huawei, Futurewei, Interdigital, QC, Nokia, ZTE] observe the performance gain of 14-70% at CSI feedback overhead A (small overhead) , for which the performance gain is 39.85%.
	+ 4 sources [Huawei, QC, Nokia, ZTE] observes the performance gain of 5-51% at CSI feedback overhead B (medium overhead)
	+ 4 source [Huawei, QC, Nokia, ZTE] observes the performance gain of 6-32% at CSI feedback overhead C (large overhead)

The following boxchart shows the median, 0.75 quantile, 0.25 quantile, outliers, and min/max values excluding outliers, for “A” CSI payload bins for various resource utilizations.



For Max Rank 4:

* For RU <= 39%, 1 source [QC] observes performance gain of 10-23%:
	+ 1 source [QC] observes a performance gain of 22.5% at CSI feedback overhead A (small overhead).
	+ 1 source [QC] observes a performance gain of 18.9% at CSI feedback overhead B (medium overhead).
	+ 1 source [QC] observes a performance gain of 10.2% at CSI feedback overhead C (large overhead).
* For RU 40-69%, 1 source [QC] observes a performance gain of 33-56%
	+ 1 source [QC] observes a performance gain of 55.4% at CSI feedback overhead A (small overhead).
	+ 1 source [QC] observes a performance gain of 48.1% at CSI feedback overhead B (medium overhead).
	+ 1 source [QC] observes a performance gain of 33.1% at CSI feedback overhead C (large overhead).
* For RU > 70%, 1 source [QC] observes a performance gain of 47-79%
	+ 1 source [QC] observes a performance gain of 79% at CSI feedback overhead A (small overhead).
	+ 1 source [QC] observes a performance gain of 69.9% at CSI feedback overhead B (medium overhead).
	+ 1 source [QC] observes a performance gain of 47.2% at CSI feedback overhead C (large overhead)..

For the evaluation of temporal domain aspects **Case 2** of AI/ML based CSI compression compared to the CSI compression Case 0 *~~benchmark~~ in terms of 5% UPT under FTP*, till RAN1 #118,

For Max rank 1,

* For RU <= 39%, 1 source [Huawei] observes performance gain of 1-5%:
	+ 1 source [Huawei] observes the performance gain of 5% at CSI feedback overhead A (small overhead)
	+ 1 source [Huawei] observes the performance gain of 3% at CSI feedback overhead B (medium overhead)
	+ 1 source [Huawei] observes the performance gain of 1% at CSI feedback overhead C (large overhead)
* For RU between 40-69%, 3 sources [Huawei, Spreadtrum, QC] observes performance gain of 1-5%:
	+ 3 sources [Huawei, Spreadtrum, QC] observes the performance gain of 3.6-5% at CSI feedback overhead A (small overhead)
	+ 3 sources [Huawei, Spreadtrum, QC] observes the performance gain of 2.8-5% at CSI feedback overhead B (medium overhead)
	+ 3 sources [Huawei, Spreadtrum, QC] observes the performance gain of 1.2-4% at CSI feedback overhead C (large overhead)
* For RU > 70%, 1 source [Huawei] observes performance gain of 1-13%:
	+ 3 sources [Huawei, Spreadtrum, QC] observe the performance gain of 7-10% at CSI feedback overhead A (small overhead)
	+ 3 sources [Huawei, Spreadtrum, QC] observe the performance gain of 4-12.9% at CSI feedback overhead B (medium overhead)
	+ 3 sources [Huawei, Spreadtrum, QC] observe the performance gain of 1-6.5% at CSI feedback overhead C (large overhead)

For Max Rank 2,

* For RU <= 39%, 4 sources [Huawei, Interdigital, Nokia, ZTE] observes performance gain of 0-12%:
	+ 4 sources [Huawei, Interdigital, Nokia, ZTE] observes the performance gain of 2-12% at CSI feedback overhead A (small overhead) , for which the performance gain is 8%.
	+ 3 sources [Huawei, Nokia, ZTE] observes the performance gain of 0-5% at CSI feedback overhead B (medium overhead)
	+ 3 source [Huawei, Nokia, ZTE] observes the performance gain of 1-3% at CSI feedback overhead C (large overhead)
* For RU between 40-69%, 5 sources [Huawei, Interdigital, QC, Nokia, ZTE] observes performance gain of 0-17%:
	+ 5 sources [Huawei, Interdigital, QC, Nokia, ZTE] observes the performance gain of 4-17% at CSI feedback overhead A (small overhead) , for which the performance gain is 13.2%.
	+ 4 sources [Huawei, QC, Nokia, ZTE] observes the performance gain of 1-14% at CSI feedback overhead B (medium overhead)
	+ 4 sources [Huawei, QC, Nokia, ZTE] observes the performance gain of 0-13% at CSI feedback overhead C (large overhead)
* For RU > 70%, 6 sources [Huawei, Futurewei, Interdigital, QC, Nokia, ZTE] observes performance gain of 3-30%:
	+ 6 sources [Huawei, Futurewei, Interdigital, QC, Nokia, ZTE] observes the performance gain of 8-30% at CSI feedback overhead A (small overhead) , for which the performance gain is 13%.
	+ 4 sources [Huawei, QC, Nokia, ZTE] observes the performance gain of 3-22% at CSI feedback overhead B (medium overhead)
	+ ~~1~~ 4 source [Huawei, QC, Nokia, ZTE] observes the performance gain of 2-17% at CSI feedback overhead C (large overhead)

The following boxchart shows the median, 0.75 quantile, 0.25 quantile, outliers, and min/max values excluding outliers, for “A” CSI payload bins for various resource utilizations.



For Max Rank 4:

* For RU <= 39%, 1 source [QC] observes performance gain of 4-10.1%:
	+ 1 source [QC] observes a performance gain of 4.6% at CSI feedback overhead A (small overhead).
	+ 1 source [QC] observes a performance gain of 10.1% at CSI feedback overhead B (medium overhead).
	+ 1 source [QC] observes a performance gain of 6.8% at CSI feedback overhead C (large overhead).
* For RU 40-69%, 1 source [QC] observes a performance gain of 4-17%
	+ 1 source [QC] observes a performance gain of 16.4% at CSI feedback overhead A (small overhead).
	+ 1 source [QC] observes a performance gain of 15% at CSI feedback overhead B (medium overhead).
	+ 1 source [QC] observes a performance gain of 4.6% at CSI feedback overhead C (large overhead).
* For RU > 70%, 1 source [QC] observes a performance gain of 11-19.1%
	+ 1 source [QC] observes a performance gain of 14.3% at CSI feedback overhead A (small overhead).
	+ 1 source [QC] observes a performance gain of 19.1% at CSI feedback overhead B (medium overhead).
	+ 1 source [QC] observes a performance gain of 11.1% at CSI feedback overhead C (large overhead)..

The above results are based on the following assumptions besides the assumptions of the agreed EVM table:

* Precoding matrix (SVD output or in angle-delay domain) is used as the model input.
* Training data samples are not quantized, i.e., Float32 is used/represented.
* 1-on-1 joint training is assumed.
* The performance metric is UPT for Max rank 1, Max rank 2, or Max rank 4.

Observation on Full buffer performance Case 2

For the evaluation of temporal domain aspects **Case 2** of AI/ML based CSI compression compared to the non-AI/ML *benchmark, in terms of mean UPT under full buffer*, till RAN1 #118,

* For Max Rank 1, 5 sources [Huawei, QC, Oppo, Xiaomi and Vivo] observe performance gains of 0-25%
	+ 5 sources [Huawei, QC, Oppo, Xiaomi and Vivo] observe performance gains of 0-25% at CSI feedback overhead A (small overhead)
	+ 5 source [Huawei, QC, Oppo, Xiaomi and Vivo] observes performance gains of 0-20% at CSI feedback overhead B (medium overhead)
	+ 5 source [Huawei, QC, Oppo, Xiaomi and Vivo] observes performance gains of 0-18% at CSI feedback overhead C (large overhead)
* For Max Rank 2, 7 sources [Huawei, Fujitsu, Xiaomi, QC, Vivo, Nokia, ZTE] observe performance gains of 1-30%
	+ 7 sources [Huawei, Fujitsu, Xiaomi, QC, Vivo, Nokia, ZTE] observe performance gains of 6-30% at CSI feedback overhead A (small overhead) , for which the performance gain is 16.6%.
	+ 7 sources [Huawei, Fujitsu, Xiaomi, QC, Vivo, Nokia, ZTE] observe performance gains of 3-23% at CSI feedback overhead B (medium overhead) , for which the performance gain is 9.3%.
	+ 7 sources [Huawei, Fujitsu, Xiaomi, QC, Vivo, Nokia, ZTE] observe performance gains of 2-24% at CSI feedback overhead C (large overhead) , for which the performance gain is 8.7%.
	+ The following boxchart shows the median, 0.75 quantile, 0.25 quantile, outliers, and min/max values excluding outliers, for (A, B, C) CSI payload bins. The source, ZTE, has 2 submissions, hence the number of points is higher than 7 in some cases.



Figure 1: Mean Throughput gains over Benchmark 1, for Bins A, B and C, for Max Rank = 2.

* For Max Rank 4, 2 sources [QC, Samsung] observe performance gains of 9-16%
	+ 2 sources [QC, Samsung] observe performance gains of 9-16% at CSI overhead A (small overhead)
	+ 1 source [QC] observes performance gains of 14.3% at CSI overhead B (medium overhead)
	+ 1 source [QC] observes performance gains of 9.2% at CSI overhead C (large overhead)

For the evaluation of temporal domain aspects **Case 2** of AI/ML based CSI compression compared to the CSI compression Case 0 *~~benchmark~~, in terms of mean UPT under full buffer*, till RAN1 #117,

* For Max Rank 1, 6 sources [Huawei, IIT Kanpur, QC, Oppo, Xiaomi and Vivo] observe performance gains of 0-13%
	+ 5 sources [Huawei, QC, Oppo, Xiaomi and Vivo] observe performance gains of 0-13% at CSI feedback overhead A (small overhead)
	+ 6 sources [Huawei, IIT Kanpur, QC, Oppo, Xiaomi and Vivo] observe performance gains of 0-12% at CSI feedback overhead B (medium overhead)
	+ 2 sources [Huawei, IIT Kanpur, QC, Oppo, Xiaomi and Vivo] observe performance gains of 0-7.7% at CSI feedback overhead C (large overhead)
* For Max Rank 2, 7 sources [Huawei, Fujitsu, Xiaomi, QC, Vivo, Nokia, ZTE] observe performance gains of 1-14%
	+ 6 sources [Huawei, Fujitsu, Xiaomi, QC, Nokia, ZTE] observe performance gains of 2-14% at CSI feedback overhead A (small overhead) , for which the performance gain is 7%.
	+ 5 sources [Huawei, Xiaomi, QC, Nokia, ZTE] observe performance gains of 1-14% at CSI feedback overhead B (medium overhead) , for which the performance gain is 4.165%.
	+ 5 sources [Huawei, Xiaomi, QC, Nokia, ZTE] observe performance gains of 1-9% at CSI feedback overhead C (large overhead) , for which the performance gain is 4.07%.
	+ The following boxchart shows the median, 0.75 quantile, 0.25 quantile, outliers, and min/max values excluding outliers, for (A, B, C) CSI payload bins.



Figure 2: Mean Throughput gains over Benchmark 2, for Bins A, B and C, for Max Rank =2.

* For Max Rank 4, 1 source [QC] observes performance gains of 8-12%
	+ 1 source [QC] observes performance gains of 11.3% at CSI overhead A (small overhead)
	+ 1 source [QC] observes performance gains of 14.3% at CSI overhead B (medium overhead)
	+ 1 source [QC] observes performance gains of 9.2% at CSI overhead C (large overhead)

For the evaluation of temporal domain aspects **Case 2** of AI/ML based CSI compression compared to the non-AI/ML *benchmark in terms of 5% UPT under full buffer*, till RAN1 #118,

* For Max Rank 1, 2 sources [QC, vivo] observe performance gains of 7-16%
	+ 2 sources [QC, vivo] observe performance gain of 10-15.3% at CSI feedback overhead A (small overhead)
	+ 2 sources [QC, vivo) observe performance gain of 7.5-8.1% at CSI feedback overhead B (medium overhead)
	+ 2 sources [QC, vivo] observe performance gain of 7.2-8.5% at CSI feedback overhead C (large overhead)
* For Max Rank 2, 5 sources [Fujitsu, QC, vivo, Nokia, ZTE] observes performance gains of 2-58%
	+ 5 sources [Fujitsu, QC, vivo, Nokia, ZTE] observe performance gains of 12-58% at CSI feedback overhead A (small overhead) , for which the performance gain is 17%.
	+ 4 sources [QC, Vivo, Nokia, ZTE] observe performance gain of 8-12% at CSI feedback overhead B (medium overhead) , for which the performance gain is 10.05%.
	+ 4 sources [QC, Vivo, Nokia, ZTE] observe performance gain of 2-9% at CSI feedback overhead C (large overhead) , for which the performance gain is7.985 %.
	+ The following boxchart shows the median, 0.75 quantile, 0.25 quantile, outliers, and min/max values excluding outliers, for (A, B, C) CSI payload bins.



Figure 3: Edge Throughput gains over Benchmark 1, for Bins A, B and C, for Max Rank =2

* For Max Rank 4, 1 source [QC] observes performance gain of 4.3-10.2%
	+ 1 source [QC] observes performance gain of 10.2% at CSI feedback overhead A (small overhead)
	+ 1 source [QC] observes performance gain of 7.4% at CSI feedback overhead B (medium overhead)
	+ 1 source [QC] observes performance gain of 4.3% at CSI feedback overhead C (large overhead)

For the evaluation of temporal domain aspects **Case 2** of AI/ML based CSI compression compared to the CSI compression Case 0 *~~benchmark~~ in terms of 5% UPT under full buffer*, till RAN1 #118,

* For Max Rank 1, 2 sources [IIT Kanpur, QC] observe performance gains of 2-14%.
	+ 1 sources [QC] observe performance gains of 2% at CSI feedback overhead A (small overhead)
	+ 2 sources [IIT Kanpur, QC] observe performance gains of 3.4-14% at CSI feedback overhead B (medium overhead)
	+ 2 sources [IIT Kanpur, QC] observe performance gains of 2.8-4.54% at CSI feedback overhead C (large overhead)
* For Max Rank 2, 4 sources [QC, Fujitsu, Nokia, ZTE] observes performance gains of 2-35%
	+ 4 sources [QC, Fujitsu, Nokia, ZTE] observe performance gains of 3-35% at CSI feedback overhead A (small overhead) , for which the performance gain is 9.81%.
	+ 3 sources [QC, Nokia, ZTE] observe performance gains of 2-6.31% at CSI feedback overhead B (medium overhead) , for which the performance gain is 4.995%.
	+ 3 sources [QC, Nokia, ZTE] observe performance gains of 2-8% at CSI feedback overhead C (large overhead) , for which the performance gain is 4.74%.
	+ The following boxchart shows the median, 0.75 quantile, 0.25 quantile, outliers, and min/max values excluding outliers, for (A, B, C) CSI payload bins.



Figure 4: Edge Throughput gains over Benchmark 2, for Bins A, B and C, for Max Rank =2

* For Max Rank 4, 1 source [QC] observes performance gain of 3-3.5%
	+ 1 source [QC] observes performance gain of 3.5% at CSI feedback overhead A (small overhead)
	+ 1 source [QC] observes performance gain of 3% at CSI feedback overhead B (medium overhead)
	+ 1 source [QC] observes performance gain of 3% at CSI feedback overhead C (large overhead)

The above results are based on the following assumptions besides the assumptions of the agreed EVM table:

* Precoding matrix (SVD output or in angle-delay domain) is used as the model input.
* Training data samples are not quantized, i.e., Float32 is used/represented.
* 1-on-1 joint training is assumed.
* The performance metric is throughput for Max rank 1, Max rank 2, or Max rank 4.

Observation on CSI feedback reduction Case 2

For the evaluation of temporal domain aspects **Case 2** of AI/ML based CSI compression, compared to the non-AI/ML benchmark, in terms of CSI feedback reduction, till RAN1 #118,

- For Max rank = 1,

- For CSI feedback overhead A (small overhead),

* 1 source [QC] observes CSI feedback reduction of 67% for full buffer;
* 1 source [QC] observes CSI feedback reduction of 17% for FTP traffic at RU <= 39%
* 1 source [QC] observes CSI feedback reduction of 73% for FTP traffic at RU of 40-69%
* 1 source [QC] observes CSI feedback reduction of 55% for FTP traffic at RU >=70%

- For CSI feedback overhead B (medium overhead),

* 1 source [QC] observes CSI feedback reduction of 67% for full buffer;
* 1 source [QC] observes CSI feedback reduction of 18% for FTP traffic at RU <= 39%
* 1 source [QC] observes CSI feedback reduction of 73% for FTP traffic at RU of 40-69%
* 1 source [QC] observes CSI feedback reduction of 55% for FTP traffic at RU >=70%

- For CSI feedback overhead C (large overhead),

* 2 sources [QC, Huawei] observes CSI feedback reduction of 68-75% for full buffer;
* 2 sources [QC, Huawei] observe the CSI feedback reduction of 47-74% for FTP traffic at RU <= 39%
* 2 sources [QC, Huawei] observes CSI feedback reduction of 78-80% for FTP traffic at RU of 40-69%
* 2 sources [QC, Huawei] observes CSI feedback reduction of 52-73% for FTP traffic at RU >=70%;

- For Max rank = 2,

- For CSI feedback overhead A (small overhead),

* 2 sources [QC, Nokia] observes CSI feedback reduction of 76-80% for full buffer;
* 1 source [QC] observes CSI feedback reduction of 38% for FTP traffic at RU <= 39%
* 1 source [QC] observes CSI feedback reduction of 83% for FTP traffic at RU of 40-69%
* 2 sources [QC, Futurewei] observe CSI feedback reduction of 69-92% at RU >= 70%

- For CSI feedback overhead B (medium overhead),

* 2 sources [QC, Nokia] observes CSI-feedback reduction of 73-80% for full buffer;
* 1 source [QC] observes CSI feedback reduction of 54% for FTP traffic at RU <= 39%
* 1 source [QC] observes CSI feedback reduction of 76% for FTP traffic at RU of 40-69%
* 1 source [QC] observe CSI feedback reduction of 67% at RU >= 70%

- For CSI feedback overhead C (large overhead),

* 3 sources [Huawei, QC, Nokia] observe the CSI feedback reduction of 70-80% for full buffer;
* 2 sources [Huawei, QC] observes the CSI feedback reduction of 5-53% for FTP traffic at RU <= 39%
* 2 sources [Huawei, QC] observes the CSI feedback reduction of 60-62% for FTP traffic at RU of 40-69%
* 2 sources [Huawei, QC] observes the CSI feedback reduction of 54-70% for FTP traffic at RU >= 70%

 - For Max rank = 4,

. - For CSI feedback overhead A (small overhead),

* 1 source [QC] observes CSI feedback reduction of 70% for full buffer;
* 1 source [QC] observes CSI feedback reduction of 73% for FTP traffic at RU <= 39%
* TBD CSI feedback reduction for FTP traffic at RU of 40-69%
* 1 sources [QC] observe CSI feedback reduction of 87% at RU >= 70%

- For CSI feedback overhead B (medium overhead),

* 1 source [QC] observes CSI-feedback reduction of 68% for full buffer;
* 1 source [QC] observes CSI feedback reduction of 55% for FTP traffic at RU <= 39%
* 1 source [QC] observes CSI feedback reduction of 83% for FTP traffic at RU of 40-69%
* 1 source [QC] observe CSI feedback reduction of 80% at RU >= 70%

- For CSI feedback overhead C (large overhead),

* 1 source [QC] observe the CSI feedback reduction of 66% for full buffer;
* 1 source [QC] observes the CSI feedback reduction of 3% for FTP traffic at RU <= 39%
* 1 source [QC] observes the CSI feedback reduction of 67% for FTP traffic at RU of 40-69%
* 1 source [QC] observes the CSI feedback reduction of 69% for FTP traffic at RU >= 70%

For the evaluation of temporal domain aspects **Case 2** of AI/ML based CSI compression, compared to the CSI compression Case 0 *~~benchmark~~*, in terms of CSI feedback reduction, till RAN1 #117,

- For Max rank = 1,

- For CSI feedback overhead A (small overhead),

* 1 source [QC] observes CSI feedback reduction of 40% for full buffer;
* TBD CSI feedback reduction for FTP traffic at RU <= 39%
* 1 source [QC] observes CSI feedback reduction of 28% at RU of 40-69%
* 1 source [QC] observes CSI feedback reduction of 23% at RU >=70%

- For CSI feedback overhead B (medium overhead),

* 1 source [QC] observes CSI-feedback reduction of 51% for full buffer;
* TBD CSI feedback reduction for FTP traffic at RU <= 39%
* 1 source [QC] observes CSI feedback reduction of 31% at RU of 40-69%
* 1 source [QC] observes CSI feedback reduction of 41% at RU >=70%

- For CSI feedback overhead C (large overhead),

* 1 source [QC] observes CSI feedback reduction of 34% for full buffer;
* TBD CSI feedback reduction for FTP traffic at RU <= 39%
* 1 source [QC] observes CSI feedback reduction of 17% at RU of 40-69%
* 1 source [QC] observes CSI feedback reduction of 29% at RU >=70%;

- For Max rank = 2,

- For CSI feedback overhead A (small overhead),

* 2 sources [QC, Nokia] observe CSI feedback reduction of 45-50%;
* TBD CSI feedback reduction for FTP traffic at RU <= 39%
* 1 source [QC] observes CSI feedback reduction of 39% at RU of 40-69%
* 1 source [QC] observes CSI feedback reduction of 42% at RU >= 70%

- For CSI feedback overhead B (medium overhead),

* 2 sources [QC, Nokia] observes CSI-feedback reduction of 50-56% for full buffer;
* TBD CSI feedback reduction for FTP traffic at RU <= 39%
* 1 source [QC] observes CSI feedback reduction of 45% at RU of 40-69%
* 1 source [QC] observes CSI feedback reduction of 45% at RU >= 70%

- For CSI feedback overhead C (large overhead),

* 3 sources [Huawei, QC, Nokia] observe CSI feedback reduction of 50-60% for full buffer,
* TBD CSI feedback reduction for FTP traffic at RU <= 39%
* 1 source [QC] observes CSI feedback reduction of 39% at RU of 40-69%
* 1 source [QC] observes CSI feedback reduction of 32% at RU >= 70%;

 - For Max rank = 4,

. - For CSI feedback overhead A (small overhead),

* 1 source [QC] observes CSI feedback reduction of 55% for full buffer;
* 1 source [QC] observes CSI feedback reduction of -3% for FTP traffic at RU <= 39%
* 1 source [QC] observes CSI feedback reductionof 45% for FTP traffic at RU of 40-69%
* 1 sources [QC] observe CSI feedback reduction of 50% at RU >= 70%

- For CSI feedback overhead B (medium overhead),

* 1 source [QC] observes CSI-feedback reduction of 59% for full buffer;
* 1 source [QC] observes CSI feedback reduction of 6% for FTP traffic at RU <= 39%
* 1 source [QC] observes CSI feedback reduction of 44% for FTP traffic at RU of 40-69%
* 1 source [QC] observe CSI feedback reduction of 55% at RU >= 70%

- For CSI feedback overhead C (large overhead),

* 1 source [QC] observe the CSI feedback reduction of 64% for full buffer;
* 1 source [QC] observes the CSI feedback reduction of 22% for FTP traffic at RU <= 39%
* 1 source [QC] observes the CSI feedback reduction of 49% for FTP traffic at RU of 40-69%
* 1 source [QC] observes the CSI feedback reduction of 54% for FTP traffic at RU >= 70%

The above results are based on the following assumptions besides the assumptions of the agreed EVM table:

* Precoding matrix (SVD output or in angle-delay domain) is used as the model input.
* Training data samples are not quantized, i.e., Float32 is used/represented.
* 1-on-1 joint training is assumed.
* The performance metric is CSI feedback overhead reduction for Max rank 1, Max rank 2, or Max rank 4

Observation on SGCS performance Case 3

For the evaluation of temporal domain aspects **Case 3** of AI/ML based CSI compression compared to the *benchmark in terms of SGCS*, for the mixed scenario of 80% indoor and 20% outdoor users:

For Layer 1,

- 6 sources [oppo, vivo, QC, Fujitsu, ZTE, DOCOMO] observe the performance gain of 1.37-28% at CSI payload X (small payload), for which the median SGCS gain is 6.95%

- 4 sources [CATT, ZTE, QC, DOCOMO] observes the performance gain of 3.9-22% at CSI payload Y (medium payload), for which the median SGCS gain is 11.05%.

- ~~4 5~~ sources [CATT, ZTE, QC, DOCOMO, MediaTek] observes the performance gain of 1.37-21% at CSI payload Z (large payload), for which the median SGCS gain is ~~8.2~~ 2.4%.

- The following boxchart shows the median, 0.75 quantile, 0.25 quantile, outliers, and min/max values excluding outliers, for (X, Y, Z) CSI payload bins.



For Layer 2,

- 4 sources [ZTE, QC, vivo, DOCOMO] observes the performance gain of 8.6-47% at CSI payload X (small payload);

- 3 sources [ZTE, QC, DOCOMO] observes the performance gain of 4.3-40% at CSI payload Y (medium payload);

- 3 sources [ZTE, QC, DOCOMO] observes the performance gain of 3.61-38% at CSI payload Z (large payload).

For Layer 3,

- 1 source [QC] observes the performance gain of 79.7% at CSI payload X (small payload);

- 1 source [QC] observes the performance gain of 28.9% at CSI payload Y (medium payload);

- 1 source [QC] observes the performance gain of 37.7% at CSI payload Z (large payload).

For Layer 4,

- 1 source [QC] observes the performance gain of 98.5% at CSI payload X (small payload);

- 1 source [QC] observes the performance gain of 33.6% at CSI payload Y (medium payload);

- 1 source [QC] observes the performance gain of 42.2% at CSI payload Z (large payload).

For the evaluation of temporal domain aspects **Case 3** of AI/ML based CSI compression compared to the *benchmark in terms of SGCS*, for the scenario of 100% outdoor users:

For Layer 1,

- 5 sources [Fujitsu, ZTE, Ericsson, Xiaomi, InterDigital] observe the performance gain of -4 to 39.76% at CSI payload X (small payload), for which the median SGCS gain is 7.44%;

- 4 sources [~~Samsung~~, ZTE, Ericsson, CMCC, Xiaomi] observe the performance gain of 1.03- 20.84% at CSI payload Y (medium payload), for which the median SGCS gain is 5.99%;

- 4 sources [~~Samsung,~~ ZTE, Ericsson, Xiaomi, InterDigital] observe the performance gain of 3.49-24.08% at CSI payload Z (large payload), for which the median SGCS gain is 5.6%.

- The following boxchart shows the median, 0.75 quantile, 0.25 quantile, outliers, and min/max values excluding outliers, for (X,Y,Z) CSI payload bins.



For Layer 2,

- 1 source [ZTE] observes the performance gain of 7.6% at CSI payload X (small payload);

- 1 source [ZTE] observes the performance gain of 6.3% at CSI payload Y (medium payload);

- 1 source [ZTE] observes the performance gain of 4.7% at CSI payload Z (large payload).

~~The description of the above boxcharts is as follows:~~

* ~~The line inside of each box is the median~~
* ~~The top and bottom box edges represent the 0.75 quantile (upper quartile) and 0.25 quantile (lower quartile), respectively.~~
* ~~Interquartile range (IQR) is the distance between the top and bottom box edges, which is used to determine the outliers (values more than 1.5 IQR away from the box edge).~~
* ~~Outliers are represented by the ‘o’ marks.~~
* ~~The whiskers represent the minimum and maximum values excluding outliers~~

Observation on FTP traffic performance Case 3

For the evaluation of temporal domain aspects **Case 3** of AI/ML based CSI compression compared to the *benchmark in terms of mean UPT* *under FTP* traffic, till RAN1 #118, for the mixed scenario of 80% indoor and 20% outdoor users:

For Max Rank 1,

* For RU <= 39%, 1 sources [CATT] observes performance gain of 0-1.3%:
	+ TBD performance gain at CSI feedback overhead A (small overhead)
	+ 1 source [CATT] observes a performance gain of 0.86% at CSI feedback overhead B (medium overhead)
	+ 1 source [CATT] observes performance gain if 1.28% at CSI feedback overhead C (large overhead)
* For RU 40-69%, 1 sources [QC] observed performance gain of 0.6-4.5%
	+ 1 sources [QC] observe performance gain of 4.5% at CSI feedback overhead A (small overhead)
	+ 1 sources [QC] observe performance gain of 0.6% at CSI feedback overhead B (medium overhead)
	+ 1 sources [QC] observe performance gain of 1.5% at CSI feedback overhead C (large overhead)
* For RU > 70%, 1 sources [QC] observes performance gain of 0.9-3.9%
	+ 1 sources [QC] observed performance gain of 2.7% at CSI feedback overhead A (small overhead)
	+ 1 sources [QC] observes performance gain of 0.9% at CSI feedback overhead B (medium overhead)
	+ 1 sources [QC] observes performance gain of 3.9% at CSI feedback overhead C (large overhead)

For Max Rank 2,

* For RU <= 39%, 1 sources [NTT Docomo] observes performance gain of 4-6%:
	+ 1 source [NTT Docomo] observes performance gain of 6.3% at CSI feedback overhead A (small overhead)
	+ 1 source [NTT Docomo] observes a performance gain of 4.2% at CSI feedback overhead B (medium overhead)
	+ 1 source [NTT Docomo] observes performance gain if 4.2% at CSI feedback overhead C (large overhead)
* For RU 40-69%, 2 sources [QC, NTT Docomo] observed performance gain of 1-12%
	+ 2 sources [QC, NTT Docomo] observe performance gain of 2-12% at CSI feedback overhead A (small overhead)
	+ 2 sources [QC, NTT Docomo] observe performance gain of 1-9.4% at CSI feedback overhead B (medium overhead)
	+ 2 sources [QC, NTT Docomo] observe performance gain of 1.6-7.1% at CSI feedback overhead C (large overhead)
* For RU > 70%, 2 sources [QC, NTT Docomo] observes performance gain of -0.2% to 22.1%
	+ 2 sources [QC, NTT Docomo] observed performance gain of -0.1% to 22.1% at CSI feedback overhead A (small overhead)
	+ 2 sources [QC, NTT Docomo] observes performance gain of -0.2% to 16.1% at CSI feedback overhead B (medium overhead)
	+ 2 sources [QC, NTT Docomo] observes performance gain of 4.3-15.2% at CSI feedback overhead C (large overhead)

For Max Rank 4,

* For RU <= 39%, 1 sources [QC] observes performance gain of 0.4-4.1%:
	+ 1 source [QC] observes performance gain of 4.1% at CSI feedback overhead A (small overhead)
	+ 1 source [QC] observes a performance gain of 0.4% at CSI feedback overhead B (medium overhead)
	+ 1 source [QC] observes performance gain if 0.5% at CSI feedback overhead C (large overhead)
* For RU 40-69%, 1 sources [QC] observed performance gain of 2-5.6%
	+ 1 sources [QC] observe performance gain of 5.6% at CSI feedback overhead A (small overhead)
	+ 1 sources [QC] observe performance gain of 2% at CSI feedback overhead B (medium overhead)
	+ 1 sources [QC] observe performance gain of 3.2% at CSI feedback overhead C (large overhead)
* For RU > 70%, 2 sources [QC] observes performance gain of 5.9% to 7.6%
	+ 1 sources [QC] observed performance gain of 7% at CSI feedback overhead A (small overhead)
	+ 1 sources [QC] observes performance gain of 5.9% at CSI feedback overhead B (medium overhead)
	+ 1 sources [QC] observes performance gain of 7.6% at CSI feedback overhead C (large overhead)

For the evaluation of temporal domain aspects **Case 3** of AI/ML based CSI compression compared to the *benchmark in terms of 5% UPT under FTP* traffic, till RAN1 #118, for the mixed scenario of 80% indoor and 20% outdoor users:

For Max Rank 1,

* For RU <= 39%, 1 sources [CATT] observes performance gain of -0.6% to 9.3%%:
	+ TBD performance gain at CSI feedback overhead A (small overhead)
	+ 1 source [CATT] observes a performance gain of -0.6% at CSI feedback overhead B (medium overhead)
	+ 1 source [CATT] observes performance gain if 9.3% at CSI feedback overhead C (large overhead)
* For RU 40-69%, 1 sources [QC] observed performance gain of 3.9-11.2%
	+ 1 sources [QC] observe performance gain of 11.2% at CSI feedback overhead A (small overhead)
	+ 1 sources [QC] observe performance gain of 3.9% at CSI feedback overhead B (medium overhead)
	+ 1 sources [QC] observe performance gain of 7.5% at CSI feedback overhead C (large overhead)
* For RU > 70%, 1 sources [QC] observes performance gain of 19.8-27.3%
	+ 1 sources [QC] observed performance gain of 19.8% at CSI feedback overhead A (small overhead)
	+ 1 sources [QC] observes performance gain of 20.7% at CSI feedback overhead B (medium overhead)
	+ 1 sources [QC] observes performance gain of 27.3% at CSI feedback overhead C (large overhead)

For Max Rank 2,

* For RU <= 39%, 1 sources [NTT Docomo] observes performance gain of 12.5-20.6%:
	+ 1 source [NTT Docomo] observes performance gain of 20.6% at CSI feedback overhead A (small overhead)
	+ 1 source [NTT Docomo] observes a performance gain of 12.5% at CSI feedback overhead B (medium overhead)
	+ 1 source [NTT Docomo] observes performance gain if 12.9% at CSI feedback overhead C (large overhead)
* For RU 40-69%, 2 sources [QC, NTT Docomo] observed performance gain of 13.4-25.3%
	+ 2 sources [QC, NTT Docomo] observe performance gain of 21-25.3% at CSI feedback overhead A (small overhead)
	+ 2 sources [QC, NTT Docomo] observe performance gain of 14.6-21% at CSI feedback overhead B (medium overhead)
	+ 2 sources [QC, NTT Docomo] observe performance gain of 13.4-18.9% at CSI feedback overhead C (large overhead)
* For RU > 70%, 2 sources [QC, NTT Docomo] observes performance gain of 27% to 51%
	+ 2 sources [QC, NTT Docomo] observed performance gain of 37.2-50.7% at CSI feedback overhead A (small overhead)
	+ 2 sources [QC, NTT Docomo] observes performance gain of 29.6-34.6% at CSI feedback overhead B (medium overhead)
	+ 2 sources [QC, NTT Docomo] observes performance gain of 27.7-31% at CSI feedback overhead C (large overhead)

For Max Rank 4,

* For RU <= 39%, 1 sources [QC] observes performance gain of 13-16%:
	+ 1 source [QC] observes performance gain of 15.6% at CSI feedback overhead A (small overhead)
	+ 1 source [QC] observes a performance gain of 13.3% at CSI feedback overhead B (medium overhead)
	+ 1 source [QC] observes performance gain if 15.7% at CSI feedback overhead C (large overhead)
* For RU 40-69%, 1 sources [QC] observed performance gain of 18-29%
	+ 1 sources [QC] observe performance gain of 28.2% at CSI feedback overhead A (small overhead)
	+ 1 sources [QC] observe performance gain of 18.5% at CSI feedback overhead B (medium overhead)
	+ 1 sources [QC] observe performance gain of 21.6% at CSI feedback overhead C (large overhead)
* For RU > 70%, 2 sources [QC] observes performance gain of 25-33%
	+ 1 sources [QC] observed performance gain of 32.2% at CSI feedback overhead A (small overhead)
	+ 1 sources [QC] observes performance gain of 27.6% at CSI feedback overhead B (medium overhead)
	+ 1 sources [QC] observes performance gain of 25.4% at CSI feedback overhead C (large overhead)

For the evaluation of temporal domain aspects **Case 3** of AI/ML based CSI compression compared to the *benchmark in terms of mean UPT* *under FTP* traffic, till RAN1 #118, for the scenario of 100% outdoor users:

For Max Rank 1,

* For RU <= 39%, 1 sources [Ericsson] observes performance gain of 0-1%:
	+ 1 source [Ericsson] observes performance gain of 0% at CSI feedback overhead A (small overhead)
	+ 1 source [Ericsson] observes performance gain of 1% at CSI feedback overhead B (medium overhead)
	+ 1 source [Ericsson] observes performance gain of 1% at CSI feedback overhead C (large overhead)
* For RU of 40-69%, 1 sources [Ericsson] observes performance gain of 4-8%:
	+ 1 source [Ericsson] observes performance gain of 4% at CSI feedback overhead A (small overhead)
	+ 1 source [Ericsson] observes performance gain of 6% at CSI feedback overhead B (medium overhead)
	+ 1 source [Ericsson] observes performance gain of 8% at CSI feedback overhead C (large overhead)
* For RU >= 70%, TBD performance gains

For the evaluation of temporal domain aspects **Case 3** of AI/ML based CSI compression compared to the *benchmark in terms of 5% UPT* *under FTP* traffic, till RAN1 #118, for the scenario of 100% outdoor users:

For Max Rank 1,

* For RU <= 39%, 1 sources [Ericsson] observes performance gain of -1% to 5%:
	+ 1 source [Ericsson] observes performance gain of -1% at CSI feedback overhead A (small overhead)
	+ 1 source [Ericsson] observes performance gain of 2% at CSI feedback overhead B (medium overhead)
	+ 1 source [Ericsson] observes performance gain of 5% at CSI feedback overhead C (large overhead)
* For RU of 40-69%, 1 sources [Ericsson] observes performance gain of 8-17%:
	+ 1 source [Ericsson] observes performance gain of 8% at CSI feedback overhead A (small overhead)
	+ 1 source [Ericsson] observes performance gain of 7% at CSI feedback overhead B (medium overhead)
	+ 1 source [Ericsson] observes performance gain of 17% at CSI feedback overhead C (large overhead)
* For RU >= 70%, TBD performance gains

The above results are based on the following assumptions besides the assumptions of the agreed EVM table:

* Precoding matrix is used as the compression model input.
* Training data samples are not quantized, i.e., Float32 is used/represented.
* 1-on-1 joint training is assumed.
* Benchmark is Rel-18 doppler eT2

Observation on Full buffer performance Case 3

For the evaluation of temporal domain aspects **Case 3** of AI/ML based CSI compression compared to the *benchmark, in terms of mean UPT under full buffer*, till RAN1 #118, for the mixed scenario of 80% indoor and 20% outdoor users,

* For Max rank 1,
	+ 3 sources [QC, Xiaomi, Vivo] observes the performance gain of 1-11% at CSI feedback overhead A (small overhead);
	+ 2 sources [QC, Xiaomi] observe performance gain of 2-5% at CSI feedback overhead B (medium overhead);
	+ 2 sources [QC, Xiaomi] observe performance gain of 2-10% at CSI feedback overhead C (large overhead);
* For Max rank 2,
	+ 4 sources [QC, Vivo, Fujitsu, ZTE] observes the performance gain of 1-19% at CSI feedback overhead A (small overhead);
	+ 2 sources [QC, ZTE] observe performance gain of 5-7% at CSI feedback overhead B (medium overhead);
	+ 2 sources [QC, ZTE] observe performance gain of 1-9% at CSI feedback overhead C (large overhead);
* For Max rank 4,
	+ 1 source [QC] observes the performance gain of 7.3% at CSI feedback overhead A (small overhead);
	+ 1 source [QC] observes performance gain of 6.3% at CSI feedback overhead B (medium overhead);
	+ 1 source [QC] observes performance gain of 8.6% at CSI feedback overhead C (large overhead);

For the evaluation of temporal domain aspects **Case 3** of AI/ML based CSI compression compared to the *benchmark in terms of 5% UPT under full buffer*, till RAN1 #118, for the mixed scenario of 80% indoor and 20% outdoor users,

* For Max rank 1,
	+ 3 sources [QC, Xiaomi, Vivo] observes the performance gain of 1-12% at CSI feedback overhead A (small overhead);
	+ 2 sources [QC, Xiaomi] observe performance gain of 2-5.3% at CSI feedback overhead B (medium overhead);
	+ 2 sources [QC, Xiaomi] observe performance gain of 4-13% at CSI feedback overhead C (large overhead);
* For Max rank 2,
	+ 4 sources [QC, Vivo, Fujitsu, ZTE] observes the performance gain of 1-34% at CSI feedback overhead A (small overhead);
	+ 2 sources [QC, ZTE] observe performance gain of 2-9% at CSI feedback overhead B (medium overhead);
	+ 2 sources [QC, ZTE] observe performance gain of 1.5-9% at CSI feedback overhead C (large overhead);
* For Max rank 4,
	+ 1 source [QC] observes the performance gain of 3% at CSI feedback overhead A (small overhead);
	+ 1 source [QC] observes performance gain of 1.2% at CSI feedback overhead B (medium overhead);
	+ 1 source [QC] observes performance gain of 7.5% at CSI feedback overhead C (large overhead);

The above results are based on the following assumptions besides the assumptions of the agreed EVM table:

- Precoding matrix ~~of the current CSI~~ is used as the compression model input.

- Training data samples are not quantized, i.e., Float32 is used/represented.

- 1-on-1 joint training is assumed.

- Benchmark is Rel-18 Type II codebook.

Observation on CSI feedback reduction Case 3

For the evaluation of temporal domain aspects **Case 3** of AI/ML based CSI compression, compared to the benchmark, in terms of CSI feedback reduction, till RAN1 #118, for the mixed scenario of 80% indoor and 20% outdoor users,

* For Max rank 1,
	+ For CSI feedback overhead A (small overhead),
		- 1 source [QC] observes CSI feedback reduction of 47% for full buffer.
		- CSI feedback reduction is TBD for FTP traffic at RU <= 39%
		- 1 source [QC] observes CSI feedback reduction of 0% for FTP traffic at RU of 40-69%
		- 1 source [QC] observes CSI feedback reduction of 29% for FTP traffic at RU >= 70%
	+ For CSI feedback overhead B (medium overhead),
		- 1 source [QC] observes CSI feedback reduction of 29% for full buffer.
		- CSI feedback reduction is TBD for FTP traffic at RU <= 39%
		- 1 source [QC] observes CSI feedback reduction of 20% for FTP traffic at RU of 40-69%
		- 1 source [QC] observes CSI feedback reduction of 8% for FTP traffic at RU >= 70%
	+ For CSI feedback overhead C (large overhead),
		- 1 source [QC] observes CSI feedback reduction of 57%
		- CSI feedback reduction is TBD for FTP traffic at RU <= 39%
		- 1 source [QC] observes CSI feedback reduction of 53% for FTP traffic at RU of 40-69%
		- 1 source [QC] observes CSI feedback reduction of 47% for FTP traffic at RU >= 70%
* For Max rank 2,
	+ For CSI feedback overhead A (small overhead),
		- 1 source [QC] observes CSI feedback reduction of 39% for full buffer.
		- CSI feedback reduction is TBD for FTP traffic at RU <= 39%
		- 1 source [QC] observes CSI feedback reduction of 23% for FTP traffic at RU of 40-69%
		- 1 source [QC] observes CSI feedback reduction of 0% for FTP traffic at RU >= 70%
	+ For CSI feedback overhead B (medium overhead),
		- 1 source [QC] observes CSI feedback reduction of 37% for full buffer.
		- CSI feedback reduction is TBD for FTP traffic at RU <= 39%
		- 1 source [QC] observes CSI feedback reduction of 11% for FTP traffic at RU of 40-69%
		- 1 source [QC] observes CSI feedback reduction of 0% for FTP traffic at RU >= 70%
	+ For CSI feedback overhead C (large overhead),
		- 1 source [QC] observes CSI feedback reduction of 49%
		- CSI feedback reduction is TBD for FTP traffic at RU <= 39%
		- 1 source [QC] observes CSI feedback reduction of 20% for FTP traffic at RU of 40-69%
		- 1 source [QC] observes CSI feedback reduction of 30% for FTP traffic at RU >= 70%
* For Max rank 4,
	+ For CSI feedback overhead A (small overhead),
		- 1 source [QC] observes CSI feedback reduction of 42% for full buffer.
		- 1 source [QC] observes CSI feedback reduction of 49% for FTP traffic at RU <= 39%
		- 1 source [QC] observes CSI feedback reduction of 52% for FTP traffic at RU of 40-69%
		- 1 source [QC] observes CSI feedback reduction of 43% for FTP traffic at RU >= 70%
	+ For CSI feedback overhead B (medium overhead),
		- 1 source [QC] observes CSI feedback reduction of 40% for full buffer.
		- 1 source [QC] observes CSI feedback reduction of 5% for FTP traffic at RU <= 39%
		- 1 source [QC] observes CSI feedback reduction of 22% for FTP traffic at RU of 40-69%
		- 1 source [QC] observes CSI feedback reduction of 34% for FTP traffic at RU >= 70%
	+ For CSI feedback overhead C (large overhead),
		- 1 source [QC] observes CSI feedback reduction of 51%
		- 1 source [QC] observes CSI feedback reduction of 12% for FTP traffic at RU <= 39%
		- 1 source [QC] observes CSI feedback reduction of 25% for FTP traffic at RU of 40-69%
		- 1 source [QC] observes CSI feedback reduction of 39% for FTP traffic at RU >= 70%

The above results are based on the following assumptions besides the assumptions of the agreed EVM table:

* Precoding matrix is used as the compression model input.
* Training data samples are not quantized, i.e., Float32 is used/represented.
* 1-on-1 joint training is assumed.
* The performance metric is CSI feedback overhead reduction for Max rank 1, Max rank 2, or Max rank 4.
* Benchmark is Rel-18 doppler eT2

Observation on SGCS performance Case 4

For the evaluation of temporal domain aspects **Case 4** of AI/ML based CSI compression compared to the *benchmark in terms of SGCS*, till RAN1 #117, for the mixed scenario of 80% indoor and 20% outdoor users:

For Layer 1,

- 1 source [QC] observes the performance gain of 7.2% at CSI payload X (small payload);

- 2 sources [QC, ETRI] observe the performance gain of 3.61-4.7% at CSI payload Y (medium payload) ;

- 1 source [QC] observes the performance gain of 7.9% at CSI payload Z (large payload) .

For Layer 2,

- 1 source [QC] observes the performance gain of 22.9% at CSI payload X (small payload);

- 1 source [QC] observes the performance gain of 10.1% at CSI payload Y (medium payload) ;

- 1 source [QC] observes the performance gain of 16.6% at CSI payload Z (large payload) .

For Layer 3,

- 1 source [QC] observes the performance gain of 80.1% at CSI payload X (small payload);

- 1 source [QC] observes the performance gain of 26.2% at CSI payload Y (medium payload) ;

- 1 source [QC] observes the performance gain of 34% at CSI payload Z (large payload) .

For Layer 4,

- 1 source [QC] observes the performance gain of 104.1% at CSI payload X (small payload);

- 1 source [QC] observes the performance gain of 32.4% at CSI payload Y (medium payload) ;

- 1 source [QC] observes the performance gain of 38.2% at CSI payload Z (large payload) .

For the evaluation of temporal domain aspects **Case 4** of AI/ML based CSI compression compared to the *benchmark in terms of SGCS*, for the scenario of 100% outdoor users:

For Layer 1,

- 1 source [Apple] observes the performance gain of 10% at CSI payload X (small payload);

- ~~The performance gain at CSI payload Y (medium payload) is TBD;~~

~~- The performance gain at CSI payload Z (large payload) is TBD.~~

Observation on SGCS performance Case 5

For the evaluation of temporal domain aspects **Case 5** of AI/ML based CSI compression compared to the non-AI/ML *benchmark in terms of SGCS*,

* For Layer 1,
	+ 2 sources [Fujitsu, OPPO] observe performance gain of 10.22-10.9% at CSI payload X (small payload)
	+ ~~Performance gain at CSI payload Y (medium payload) is TBD~~
	+ ~~Performance gain at CSI payload Z (large payload) is TBD~~

For the evaluation of temporal domain aspects **Case 5** of AI/ML based CSI compression compared to the CSI compression Case 0 *in terms of SGCS*,

* For Layer 1,
	+ 2 sources [Fujitsu, OPPO] observe performance gain of 1.7-6.3% at CSI payload X (small payload)
	+ 1 source [IIT Kanpur] observes performance gain of 39.5% at CSI payload Y (medium payload)
	+ 1 source [IIT Kanpur] observes performance gain of 6.62% at CSI payload Z (large payload)

The above results are based on the following assumptions besides the assumptions of the agreed EVM table:

* Precoding matrix is used as the model input.
* Training data samples are not quantized, i.e., Float32 is used/represented.
* 1-on-1 joint training is assumed.
* The performance metric is SGCS for Layer 1 of Max rank 1 or Layer 1/2 of Max rank 2.
* CSI payload X is ≤ 80/α bits; CSI payload Y is (100 - 140 )/α bits; CSI payload Z is ≥ 230/α bits; where X, Y, Z are applicable per layer, where alpha = 1 for rank = 1/2 and alpha = 2 for rank = 3/4.
* Benchmark is Rel-16 Type II codebook.

Observation on SGCS performance multi-vendor training

For the multi-vendor evaluation of temporal domain aspects **Case 1/2/3/4/5** of AI/ML based CSI compression assuming Type 3 NW first separate training for the pairing of 1 NW to 1 UE (Case 1-NW first), by comparing the performance with 1-to-1 joint training baseline:

For Case 2

- 1 source [Huawei] observes the performance gain of 0.1% at CSI payload X (small payload);

- 2 sources [QC, Huawei] observe the performance gain of -0.86% to -0.12% at CSI payload Y (medium payload) ;

- 2 sources [QC, Huawei] observe the performance gain of -0.74% to -0.02% at CSI payload Z (large payload) .

For Case 3

- 1 source [QC] observes the performance gain of -0.65% to -0.53% at CSI payload Y (medium payload) ;

- 1 source [QC] observes the performance gain of -0.40% to -0.33% at CSI payload Z (large payload).

The above results are based on the following assumptions besides the assumptions of the agreed EVM table:

* Precoding matrix is used as the model input.
* Training data samples are not quantized, i.e., Float32 is used/represented.
* Separate training with dataset sharing, where shared output of CSI generation part is before [Huawei] or after [QC] quantization
* Quantization/dequantization method/parameters between NW side and UE side are aligned.

Observation on SGCS performance and complexity of localized models Option 1

For the evaluation of AI/ML based CSI compression using localized models (Option 1), compared to the non-AI/ML *benchmark in terms of SGCS*, till RAN1 #118

For Layer 1,

* + ~~2~~ 3 sources [ZTE, vivo, Intel] observe the performance gain of ~~15~~ 4.5-28% over benchmark, at CSI payload X (small payload), ~~using TSF compression~~
	+ 2 sources [ZTE, vivo] observes the performance gain of 13-15 % at CSI payload Y (medium payload), ~~using TSF compression~~
	+ 2 sources [ZTE, vivo] observes the performance gain of 8-11% at CSI payload Z (large payload), ~~using TSF compression~~

For Layer 2,

* + 2 sources [ZTE, vivo] observes a performance gain of 15-24% over benchmark, at CSI payload of X (small payload)
	+ 2 sources [ZTE, vivo] observes a performance gain of 16-21% over benchmark, at CSI payload of Y (medium payload)
	+ 2 sources [ZTE, vivo] observes a performance gain of 9-18% over benchmark, at CSI payload of Z (large payload)

For the evaluation of AI/ML based CSI compression using localized models (Option 1), compared to the AI/ML based CSI compression using global models *~~benchmark~~ in terms of SGCS*, where the localized models have the same complexity as the global models, till RAN1 #118,

For Layer 1,

* + ~~3~~ 4 sources [ZTE, vivo, QC, Intel] observes the performance gain of 1-11% over global model, at CSI payload X (small payload)~~, using TSF compression~~
	+ 3 sources [ZTE, vivo, QC] observes the performance gain of 4-17% at CSI payload Y (medium payload) ~~, using TSF compression~~
	+ 2 sources [ZTE, vivo] observes the performance gain of 1-7% at CSI payload Z (large payload) ~~, using TSF compression~~

For Layer 2,

* + 2 sources [ZTE, vivo] observes a performance gain of 5-14% over global model, at CSI payload of X (small payload)
	+ 2 sources [ZTE, vivo] observes a performance gain of 5-19% over global model, at CSI payload of Y (medium payload)
	+ 2 sources [ZTE, vivo] observes a performance gain of 0-13% over global model, at CSI payload of Z (large payload)

For the evaluation of AI/ML based CSI compression using localized models (Option 1), compared to the AI/ML based CSI compression using global models *~~benchmark~~ in terms of SGCS*, where the localized models have the reduced complexity as the global models, till RAN1 #118,

For Layer 1,

* + 1 source [QC] observes the performance gain of 1.6% over global model, with 27% of parameters and 1% of FLOPs as the global model, at CSI payload X (small payload) ; while 1 source [Intel] observes 0% gain over global model, with 2.8% of parameters and 2.6% of FLOPs as the global model,
	+ 1 source [QC] observes the performance gain of 11.9% over the global model, with 27% of parameters and 1% of FLOPs as the global model,  at CSI payload Y (medium payload)
	+ ~~TBD performance gain at CSI payload Z (large payload)~~

The above results are based on the following assumptions besides the assumptions of the agreed EVM table:

* Precoding matrix of the current CSI is used as the model input.
* Training data samples are not quantized, i.e., Float32 is used/represented.
* 1-on-1 joint training is assumed.
* The performance metric is SGCS for Layer 1 of Max rank 1 or Layer 1/2 of Max rank 2.
* CSI payload X is ≤ 80 bits; CSI payload Y is 100 bits - 140 bits; CSI payload Z is ≥ 230 bits; where X, Y, Z are applicable per layer.
* Benchmark is Rel-16 Type II codebook.

Observation on SGCS performance and complexity of localized models Option 2

For the evaluation of AI/ML based CSI compression using localized models (Option 2), compared to the non-AI/ML *benchmark in terms of SGCS*, till RAN1 #118

For Layer 1,

* + 3 sources [ViVo, Oppo, Intel] observe a performance gain of 4.5-10% over benchmark, at CSI payload of X (small payload)
	+ Performance gain of TBD % over benchmark, at CSI payload of Y (medium payload)
	+ Performance gain of TBD % over benchmark, at CSI payload of Y (medium payload)

For the evaluation of AI/ML based CSI compression using localized models (Option 2), compared to the AI/ML based CSI compression using global models *~~benchmark~~ in terms of SGCS*, where the localized models have the same complexity as the global models, till RAN1 #118,

For Layer 1,

* + 6 sources [Nokia, Vivo, Panasonic, Oppo, Futurewei, Intel] observes a performance gain of -2.65% to 6% over global model, at CSI payload of X (small payload)
	+ 1 source [Nokia] observes a performance gain of 0-2% over global model, at CSI payload of Y (medium payload)
	+ 1 source [Nokia] observes a performance gain of -0.5% to 3% over global model, at CSI payload of Y (medium payload)

For the evaluation of AI/ML based CSI compression using localized models (Option 2), compared to the AI/ML based CSI compression using global models *~~benchmark~~ in terms of SGCS*, where the localized models have the lower complexity as the global models, till RAN1 #118,

For Layer 1,

* + 1 sources [Intel] observes a performance gain of 0% over global model, with 2.8% of parameters and 2.6% of FLOPs as the global model, at CSI payload of X (small payload)
	+ TBD performance gain over global model, at CSI payload of Y (medium payload)
	+ TBD performance gain over global model, at CSI payload of Y (medium payload)

The above results are based on the following assumptions besides the assumptions of the agreed EVM table:

* Precoding matrix of the current CSI is used as the model input.
* Training data samples are not quantized, i.e., Float32 is used/represented.
* 1-on-1 joint training is assumed.
* The performance metric is SGCS for Layer 1 of Max rank 1 or Layer 1/2 of Max rank 2.
* CSI payload X is ≤ 80 bits; CSI payload Y is 100 bits - 140 bits; CSI payload Z is ≥ 230 bits; where X, Y, Z are applicable per layer.
* Benchmark is Rel-16 Type II codebook.

Observation on FTP performance of localized models Option 2

For the evaluation of localized models of AI/ML based CSI compression compared to the *benchmark, in terms of mean UPT under FTP traffic*, till RAN1 #118,

* For Max Rank 1,
* For RU <= 39%, 1 source [Intel] observes 0% performance gain
* For RU of 40-69%, 1 source [Intel] observes 0.1% performance gains
* For RU >= 70%, 1 source [Intel] observes 0.9% performance gains

For the evaluation of localized models of AI/ML based CSI compression compared to compression using global models, *in terms of mean UPT under FTP traffic,* where localized models have the same complexity as the *global model,* till RAN1 #118,

* For Max Rank 1,
* For RU <= 39%, 1 source [Intel] observes 0% performance gain
* For RU of 40-69%, 1 source [Intel] observes 0.2% performance gains
* For RU >= 70%, 1 source [Intel] observes 1.2% performance gains

For the evaluation of localized models of AI/ML based CSI compression compared to compression using global models, *in terms of mean UPT under FTP traffic,* where localized models have lower complexity than the *global model,* till RAN1 #118,

* For Max Rank 1,
* For RU <= 39%, 1 source [Intel] observes ~~0%~~ -0.1% performance gain over the global model, with 2.8% of parameters and 2.6% of FLOPs as the global model, at CSI feedback overhead A (small overhead).
* For RU of 40-69%, 1 source [Intel] observes ~~0.3%~~ -0.5% performance gains over the global model, with 2.8% of parameters and 2.6% of FLOPs as the global model, at CSI feedback overhead A (small overhead)
* For RU >= 70%, 1 source [Intel] observes ~~0.9%~~ -1.4% performance gains over the global model, with 2.8% of parameters and 2.6% of FLOPs as the global model, at CSI feedback overhead A (small overhead)

For the evaluation of localized models of AI/ML based CSI compression compared to the *benchmark, in terms of 5% UPT under FTP traffic*, till RAN1 #118,

* For Max Rank 1,
* For RU <= 39%, 1 source [Intel] observes -0.4% performance gain
* For RU of 40-69%, 1 source [Intel] observes 0.6% performance gains
* For RU >= 70%, 1 source [Intel] observes 0.6% performance gains

For the evaluation of localized models of AI/ML based CSI compression compared to compression using global models, *in terms of 5% UPT under FTP traffic,* where localized models have the same complexity as the *global model,* till RAN1 #118,

* For Max Rank 1,
* For RU <= 39%, 1 source [Intel] observes -0.6% performance gain
* For RU of 40-69%, 1 source [Intel] observes 0.8% performance gains
* For RU >= 70%, 1 source [Intel] observes 2.5% performance gains

For the evaluation of localized models of AI/ML based CSI compression compared to compression using global models, *in terms of 5% UPT under FTP traffic,* where localized models have lower complexity than the *global model,* till RAN1 #118,

* For Max Rank 1,
* For RU <= 39%, 1 source [Intel] observes ~~0.4%~~ 0.2% performance gain over the global model, with 2.8% of parameters and 2.6% of FLOPs as the global model, at CSI feedback overhead A (small overhead)
* For RU of 40-69%, 1 source [Intel] observes ~~2.0%~~ -2.4% performance gains over the global model, with 2.8% of parameters and 2.6% of FLOPs as the global model, at CSI feedback overhead A (small overhead)
* For RU >= 70%, 1 source [Intel] observes 3~~.8%~~ -4.1% performance gains over the global model, with 2.8% of parameters and 2.6% of FLOPs as the global model, at CSI feedback overhead A (small overhead).

Observation on Full buffer performance of localized models Option 1

For the evaluation of AI/ML based CSI compression using localized models (Option 1), compared to the *benchmark, in terms of mean UPT under full buffer*, till RAN1 #118,

* For Max Rank 1, TBD performance gains
* For Max Rank 2,
* 1 source [Vivo] observes performance gain of 10.8-15.2%,
	+ 1 source [Vivo] observes the performance gain of 15.2% at CSI feedback overhead A (small overhead);
	+ 1 source [Vivo] observes the performance gain of 10.8% at CSI feedback overhead B (medium overhead);
	+ 1 source [Vivo] observes the performance gain of 11.9% at CSI feedback overhead A (small overhead)
* For Max Rank 4, TBD performance gains.

For the evaluation of AI/ML based CSI compression using localized models (Option 2), compared to the compression using global models *~~benchmark~~ in terms of mean UPT under full buffer*, where the localized models have the same complexity as the global models, till RAN1 #118,

* For Max Rank 1, TBD performance gains
* For Max Rank 2,
* 1 source [Vivo] observes performance gain of 3.87-6.43%,
	+ 1 source [Vivo] observes the performance gain of 3.87% at CSI feedback overhead A (small overhead);
	+ 1 source [Vivo] observes the performance gain of 6.43% at CSI feedback overhead B (medium overhead);
	+ 1 source [Vivo] observes the performance gain of 6.36% at CSI feedback overhead A (small overhead)
* For Max Rank 4, TBD performance gains.

For the evaluation of AI/ML based CSI compression using localized models (Option 1), compared to the *benchmark, in terms of 5% UPT under full buffer*, till RAN1 #118,

* For Max Rank 1, TBD performance gains
* For Max Rank 2,
* 1 source [Vivo] observes performance gain of 12.8-20.5%,
	+ 1 source [Vivo] observes the performance gain of 17.6% at CSI feedback overhead A (small overhead);
	+ 1 source [Vivo] observes the performance gain of 12.8% at CSI feedback overhead B (medium overhead);
	+ 1 source [Vivo] observes the performance gain of 20.5% at CSI feedback overhead A (small overhead)
* For Max Rank 4, TBD performance gains.

For the evaluation of AI/ML based CSI compression using localized models (Option 2), compared to the compression using global models *~~benchmark~~ in terms of 5% UPT under full buffer*, where the localized models have the same complexity as the global models, till RAN1 #118,

* For Max Rank 1, TBD performance gains
* For Max Rank 2,
* 1 source [Vivo] observes performance gain of 2.52-9.83%,
	+ 1 source [Vivo] observes the performance gain of 2.52% at CSI feedback overhead A (small overhead);
	+ 1 source [Vivo] observes the performance gain of 9.83% at CSI feedback overhead B (medium overhead);
	+ 1 source [Vivo] observes the performance gain of 9.24% at CSI feedback overhead A (small overhead)
* For Max Rank 4, TBD performance gains.

Observation on Full buffer performance of localized models Option 2

For the evaluation of AI/ML based CSI compression using localized models (Option 2), compared to the *benchmark, in terms of mean UPT under full buffer*, till RAN1 #118,

* For Max Rank 1,
* 1 source [Intel] observes performance gain of 3.7%,
	+ 1 source [Intel] observes the performance gain of 3.7% at CSI feedback overhead A (small overhead);
	+ TBD performance gain at CSI feedback overhead B (medium overhead)
	+ TBD performance gain at CSI feedback overhead C (large overhead)
* For Max Rank 2, TBD performance gains
* For Max Rank 4, TBD performance gains.

For the evaluation of AI/ML based CSI compression using localized models (Option 2), compared to the compression using global models *~~benchmark~~ in terms of mean UPT under full buffer*, where the localized models have the same complexity as the global models, till RAN1 #118

* For Max Rank 1,
* 1 source [Intel] observes performance gain of 3.8%,
	+ 1 source [Intel] observes the performance gain of 3.8% at CSI feedback overhead A (small overhead);
	+ TBD performance gain at CSI feedback overhead B (medium overhead)
	+ TBD performance gain at CSI feedback overhead C (large overhead)
* For Max Rank 2, TBD performance gains
* For Max Rank 4, TBD performance gains.

For the evaluation of AI/ML based CSI compression using localized models (Option 2), compared to the compression using global models *~~benchmark~~ in terms of mean UPT under full buffer*, where the localized models have lower complexity as the global models, till RAN1 #118,

* For Max Rank 1,
* 1 source [Intel] observes performance gain of ~~2.7%~~ -1.2%,
	+ 1 source [Intel] observes the performance gain of ~~2.7%~~ -1.2% over the global model, with 2.8% of parameters and 2.6% of FLOPs as the global model, at CSI feedback overhead A (small overhead);
	+ TBD performance gain at CSI feedback overhead B (medium overhead)
	+ TBD performance gain at CSI feedback overhead C (large overhead)
* For Max Rank 2, TBD performance gains
* For Max Rank 4, TBD performance gains.

For the evaluation of localized models of AI/ML based CSI compression compared to the *benchmark, in terms of 5% edge UPT under full buffer*, till RAN1 #118,

* For Max Rank 1,
* 1 source [Intel] observes performance gain of -2.0%,
	+ 1 source [Intel] observes the performance gain of -2% at CSI feedback overhead A (small overhead);
	+ TBD performance gain at CSI feedback overhead B (medium overhead)
	+ TBD performance gain at CSI feedback overhead C (large overhead)
* For Max Rank 2, TBD performance gains
* For Max Rank 4, TBD performance gains.

For the evaluation of localized models of AI/ML based CSI compression compared compression using global models*, in terms of 5% edge UPT under full buffer*, where the localized models have same complexity as the global models, till RAN1 #118,

* For Max Rank 1,
* 1 source [Intel] observes performance gain of 2.2%,
	+ 1 source [Intel] observes the performance gain of 2.2% at CSI feedback overhead A (small overhead);
	+ TBD performance gain at CSI feedback overhead B (medium overhead)
	+ TBD performance gain at CSI feedback overhead C (large overhead)
* For Max Rank 2, TBD performance gains
* For Max Rank 4, TBD performance gains.

For the evaluation of localized models of AI/ML based CSI compression compared to compression using global models*, in terms of 5% edge UPT under full buffer*, where the localized models have lower complexity as the global models, till RAN1 #118,

* For Max Rank 1,
* 1 source [Intel] observes performance gain of ~~3.4%~~ -3%,
	+ 1 source [Intel] observes the performance gain of ~~3.4%~~ -3% over the global model, with 2.8% of parameters and 2.6% of FLOPs as the global model, at CSI feedback overhead A (small overhead);
	+ TBD performance gain at CSI feedback overhead B (medium overhead)
	+ TBD performance gain at CSI feedback overhead C (large overhead)
* For Max Rank 2, TBD performance gains
* For Max Rank 4, TBD performance gains.

Observation: Performance complexity trade-off plots

Performance vs. complexity for temporal domain Case 0

In all the plots, the x-axis is the combined complexity of the CSI generation part and the CSI reconstruction part.

Performance vs. complexity for temporal domain Case 2

In all the plots, the x-axis is the combined complexity of the CSI generation part and the CSI reconstruction part.

Performance vs. complexity for temporal domain Case 3

In all the plots, the x-axis is the combined complexity of the CSI generation part and the CSI reconstruction part.

For Case 3, the FLOPs represent the total FLOP over the time window, not the normalized FLOP over 5msec. (i.e., the FLOPs is based on FLOPs/M from the results template, not the FLOPs/M/5msec.)

# Company comments

|  |  |
| --- | --- |
| *Company* | *Comments* |
|  |  |
|  |  |