3GPP TSG-RAN Rel-18 workshop **RWS-210307**

Electronic meeting, June 28th – July 2nd 2021

Agenda Item: **4.1**

Source: **Ericsson**

Title: **Discussion on Release 18 MIMO enhancements**

Document for: **Discussion**

1 Introduction

In this document, we discuss and share our view on topics for a Rel.18 work item on further enhanced and evolved MIMO (feeMIMO).

2 Justification of Release 18

We have identified a set of enhancements which can be categorized under the main topics as indicated by the subsections below. Here we give the motivation and justification for these.

2.1 NR UL MIMO enhancements to be on par with LTE UL MIMO

The NR uplink has two waveforms that are semi-statically configured precoded (DFT-S-) and non-precoded (CP-) OFDM. DFT-S-OFDM has a relatively low peak to average power probability, which allows the UE to operate its power amplifier closer to its maximum output power (theoretically ~3 dB closer for QPSK as measured by the cubic metric) than CP-OFDM. This higher output power can be used to improve coverage, especially for higher modulation orders. It also enables more efficient PA operation, and consequently reduced current drain in the UE.

While DFT-S-OFDM has benefits, it also has drawbacks. DFT-S-OFDM scheduling is restricted to contiguous PRBs in the frequency domain, and the number of PRBs must be a multiple of 2, 3, and 5 (i.e. $N_{PRB} = 2^i 3^j 5^k$, with i, j, and k non-negative integers). Furthermore, DFT-S-OFDM produces inter-subcarrier interference in the presence of delay spread which can require equalization and/or degrade performance. CP-OFDM on the other hand allows non-contiguous resource allocation, any integer number of PRBs in the frequency domain, and generally does not require equalization.

Comparing DFT-S-OFDM and CP-OFDM, we observe that their tradeoffs are a function of dynamic parameters. The benefit of non-contiguous frequency domain resource allocation or unrestricted numbers of PRBs may be more desirable to efficiently pack UEs in the frequency domain while reaping the benefits of frequency selective scheduling. Such benefits accrue faster at higher cell loads. If the UE is in a relatively flat channel and where frequency domain scheduling constraints are not a concern, using DFT-S-OFDM rather than CP-OFDM could allow the UE to operate at a higher MCS state because it will be able to transmit closer to its maximum rated power. Similarly, if the UE experiences a deep fade or sudden blockage of a UE antenna, being able to transmit at the highest power level can be beneficial, and switching to DFT-S-OFDM could improve performance.

Observation 1 **Tradeoffs between DFT-S-OFDM and CP-OFDM operation are often a function of dynamic parameters, such as cell load, scheduling / link adaptation, fading, and/or antenna blockage.**

Proposal 1 **Specify faster than RRC switching of UL waveform.**

The constraints of using the DFT-S-OFDM waveform for UL-MIMO in NR are even more restrictive than for non-MIMO, since NR supports only a single layer for DFT-S-OFDM. In LTE, DFT-SOFDM is specified with up to four layer transmission and it is straightforward to enhance NR to be on par with LTE uplink spectral efficiency for this single carrier waveform case.

Currently, DFT-S-OFDM is not used in the upper QPSK MCS range due to lack of 2 layer transmission. In general, the benefit for cell edge / coverage scenarios may not be obvious. In practice it turns out that rank 2 or higher transmission can be quite common in a cell, and that multilayer transmission can be a mechanism to deliver higher power especially for non-coherent UL MIMO UEs.

Figure 1 below shows a histogram of the UL MIMO rank in a cell when the gNB has 4 or 32 Rx antennas. Rel-15 non-coherent UL MIMO transmission is used, and FTP model 1 traffic is used. Resource utilization is roughly 40%. The details of the simulation setup and further discussion can be found in R1-2008419. It can be seen that very few UEs transmit only rank 1. In the 4 Rx case, less than 1% of the UEs transmit rank 1, while for 32 gNB Rx antennas, rank 2 is always used. One major reason for the use of high rank is that non-coherent UEs gain 3 dB more power by transmitting two layers.

Therefore, the cubic metric gain of ~ 3dB from DFT-S-OFDM can be reaped over the vast majority of the cell, instead of being constrained toward the center of the cell.

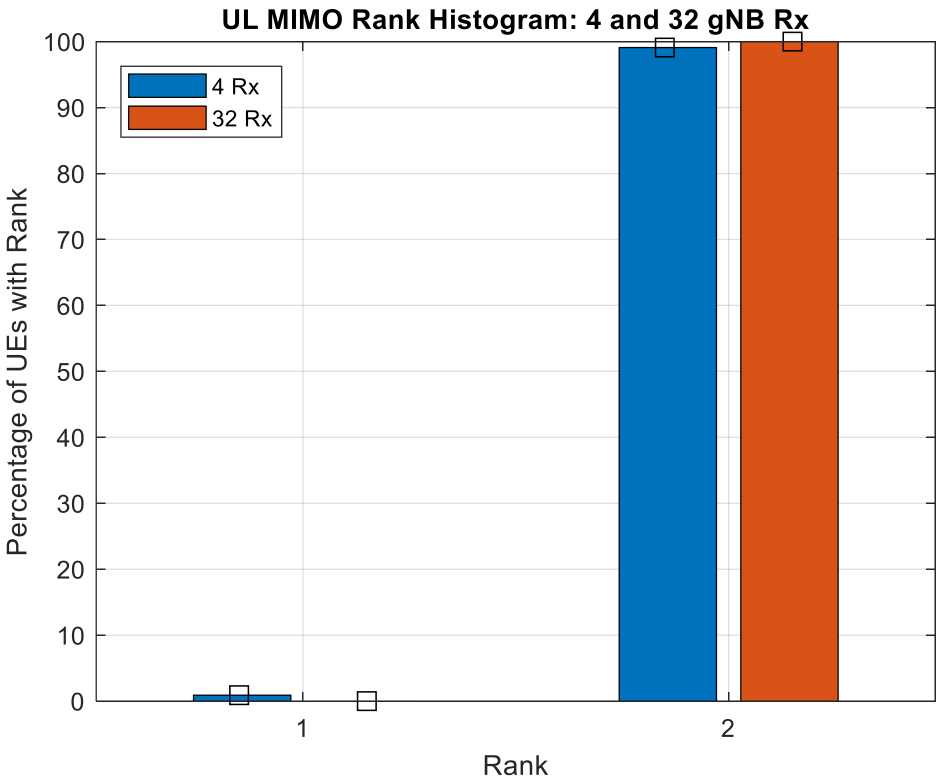

Figure 1. UL MIMO rank histograms for 4 and 32 Rx gNB

Proposal 2 **Specify multi-layer PUSCH for DFT-S-OFDM.**

Another UL MIMO enhancement is targeting UL MU-MIMO. It is observed in our preliminary analysis that configuring double-symbol DMRS creates too large DMRS overhead (4 out of 14 symbols in the slot is DMRS) and actually reduces the throughput compared to a single DMRS symbol.

This is the case even though non-orthogonal DMRS is used (a single symbol Type 1 DMRS has only four orthogonal DMRS ports) to maintain the same number of received MU-MIMO layers in gNB, i.e. to maintain the large number of DMRS ports needed for UL MU-MIMO. It will thus benefit UL performance in the MU-MIMO case to specify an increase the number of orthogonal DMRS ports in a single DMRS symbol.

It is also noted that MU-MIMO with a large number of co-scheduled users is mainly considered in deployments with low delay spread, hence, to increase the number of orthogonal ports using cyclic shifts or sparser comb seems to be possible. Aspects of non-backward compatibility needs to be considered in the work.

Proposal 3 **Specify DMRS capacity enhancements for Type 1 DMRS to at least double the number of orthogonal DMRS ports in one OFDM symbol.**

2.2 More precise and faster CSI reporting

Link adaptation in NR and LTE is rather slow and inaccurate. Often an open loop link adaptation (OLLA) algorithm is needed in the gNB to adjust the CSI reported from the UE to make a better MCS selection for PDSCH. This is due to bursty interference, coarse CQI tables, and when short infrequent bursts of packets are used, which doesn't give a chance for OLLA to converge.

In addition, current CSI feedback may be simplified and may not take into account all benefits of the advanced receiver, which also increase the need for OLLA. Likewise, current CSI doesn't take into account receiver impairments, DMRS channel estimation and interference estimation errors, frequency offsets and also phase noise impairments.

Also, current CSI may be simplified and codebook based, and the true PDSCH transmission may use another precoder, have different inter-MIMO-layer interference characteristics, and the UE may use different methods to compute SINR for CQI and SINR for PDSCH demodulation.

A feature that has been discussed recently would help this situation. It is based on introducing additional CSI feedback which is determined from the actual received PDSCH quality of a previous scheduling. Such feedback can be tagged on along with the HARQ-ACK. When a gNB receives this new information, it can directly adjust the MCS for the next, similar, transmission, without the need for OLLA adjustments. Hence, the UE informs the gNB of a CQI computed post-decoding of PDSCH.

Similar ideas have been floated, for example DMRS based CQI and soft-HARQ-ACK. The DMRS based CQI is yet another approximation for the "true" CQI for the received PDSCH. The soft-HARQ-ACK is on the other hand a better representation of such.

Note that in the Rel-17 IIoT/URLLC work item, there is a closely related proposal for 'delta-CQI/MCS' that is currently being studied. However, as discussed in our paper [1], the 'delta-CQI/MCS' scheme is still not agreed to be specified in Rel-17. If the 'delta-CQI/MCS' scheme is not specified in Rel-17 IIoT/URLLC work item, then we suggest that Rel-18 MIMO work item contains a study of different techniques and their benefits of post-decoding CQI.

Proposal 4 **Study, and if beneficial, specify post-decoding CSI feedback for PDSCH (pending outcome of Rel-17 IIoT/URLLC discussion on 'delta-CQI/MCS').**

Another enhancement related to CSI feedback is a feature designed to acquire a fast CSI feedback during handover and initial access so that the UE can be scheduled with a better MCS directly. With link adaptation, the network (NW) adjusts the modulation and coding schemes to match the instantaneous channel conditions. In DL, the adaptation is based on CSI reports from the UE. The UE performs measurements on CSI-RS and send the measurement to the NW. The UE cannot perform any measurements or send any CSI report to the network until the RRC configuration has been completed. Before that, the network would have to assume that the UE is at the cell border and is forced to use quite robust modulation and coding scheme.

The four-step random access procedure is depicted in Figure 2.

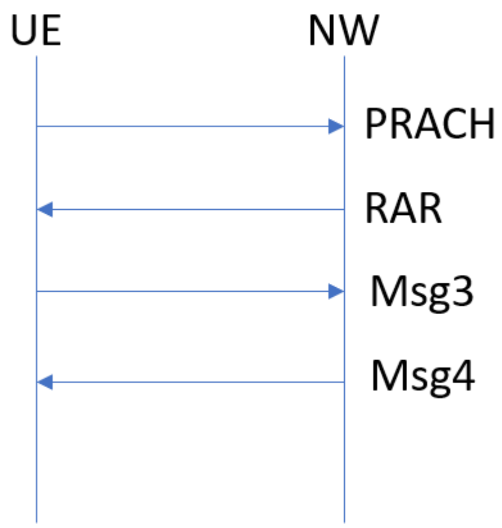

Figure 2: Four step random access.

In RAR, the NW may include a CSI request: in current specifications, there is a bit reserved for that purpose, as described in Table 8.2-1 in TS 38.213. However, there is no description of what that bit would be used for. The CSI request field was introduced in NB-IoT, with the purpose to aid the PDCCH link adaptation: the UE would provide the NW with a CSI report already in Msg3. This is described in TS 36.133 and TS 36.331.

We propose to introduce such CSI reporting also for NR:

Proposal 5 **Specify early CSI reporting in Msg3 triggered by the already existing CSI request bit in the RAR grant, both for contention-based and non-contention-based random access.**

2.3 Multi-TRP Enhancements for URLLC

Multi-TRP for URLLC still has limitations as there is a need to ensure improved spectral efficiency while providing high reliability; CSI feedback can be inaccurate as current CSI does not consider Multi-TRP URLLC schemes; URLLC CSI triggering is currently inflexible as it cannot be triggered when a high latency CSI computation is ongoing.

In NR Rel-16 specifications, when multi-DCI multi-TRP is configured in a CC, single-DCI based multi-TRP reliability schemes for URLLC cannot be used. As a result, diversity transmission over multi-TRP cannot be achieved. For intra-UE multiplexing use cases in FR2, the whole bandwidth would be used for either eMBB or URLLC data.

When multi-DCI is configured, eMBB and URLLC packets would have to be scheduled in TDM fashion which results in very low spectrum efficiency when scheduling small URLLC packets. To support high reliability with improved spectral efficiency for intra-UE multiplexing (with eMBB and URLLC traffics), introducing mixed mode single-DCI and multi-DCI multi-TRP is beneficial. Enabling mixed mode single-DCI and multi-DCI multi-TRP would allow eMBB data to be scheduled at the same time with URLLC data with multi-TRP repetition (i.e., eMBB scheduled in one CORESET Pool and URLLC scheduled in the other) as shown in Figure 3. Hence, we make the following proposal for NR Rel-18:

Proposal 6 **Specify mixed mode single-DCI and multi-DCI multi-TRP to support high reliability with improved spectral efficiency for intra-UE multiplexing.**

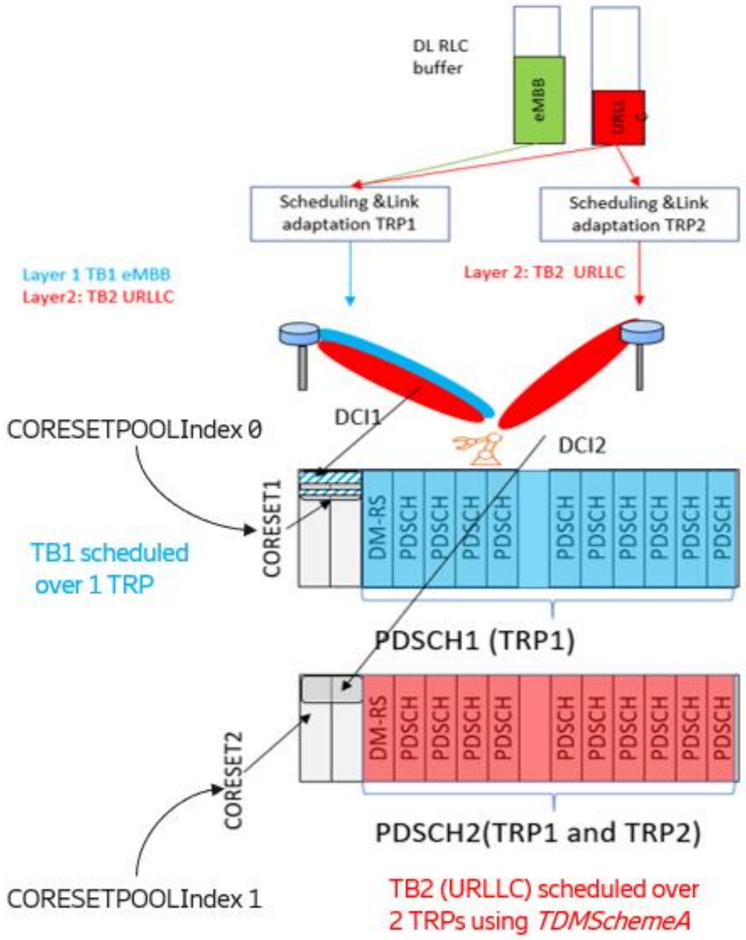

Figure 3. An example of mixed mode single-DCI and multi-DCI multi-TRP

In NR Rel-16, there is no CSI reporting scheme optimized for the multi-TRP URLLC transmission schemes. For example, a single PDCCH can schedule PDSCH(s) from two TRPs using same or different redundancy version. However, there is no CSI reporting scheme where the CSI reflects this fact. The best gNB can do is to configure UE with two CSI report configurations with different TCI states associated with the two TRPs. The gNB will then receive two CSIs each indicating a three-tuple (RI, PMI, CQI) and from those CSIs deduce a single rank, two pre-coders and a single MCS to be used for scheduling one of the multi-TRP URLLC schemes. This is by no means straight forward and the deduced CSI can be inaccurate. Thus, it is desirable to let the UE to report a CSI by taking the multi-TRP URLLC schemes (e.g., *FDMSchemeA*, *FDMSchemeB*, *TDMSchemeA*) being used for transmission into account. Hence, we propose the following:

Proposal 7 Specify Multi-TRP CSI enhancements for multi-TRP URLLC schemes.

The ultra-low latency CSI timing specified in NR is useful when the gNB needs to rapidly schedule URLLC data and thus acquire CSI quickly to use a correct link adaptation for the URLLC transmission. The ultra-low latency timing can only be applied when the UE is not already calculating any other CSI report, which is quite restrictive. Since the 'High Latency' CSI reports require comparably long CSI processing times, when a 'High Latency' CSI is being processed, ultra-low latency CSI cannot be triggered.

Hence, we propose introducing support for flexible URLLC CSI triggering even when a 'High Latency' CSI is being processed. One possibility is to allow URLLC CSI reports to "override" previous CSI calculations and thus allocate all CSI computation resources for the URLLC CSI report. The key benefit of introducing more flexible triggering of URLLC CSI report, is that it allows the ultra-low latency CSI timeline to be applied for more cases, which reduces the CSI latency and in turn results in more reliable URLLC scheduling.

Proposal 8 Specify flexible URLLC CSI triggering even when a 'High Latency' CSI is being processed.

2.4 Multi-TRP dynamic point selection enhancement

When performing the TRS design in Rel.15, many evaluations were made and presented in RAN1. It was observed that TRS is only useful for SNR<3 dB and 6 RB scheduling BW and for QPSK modulation. Also, our results showed that for wideband scheduling (50 RB PDSCH), the performance when performing synchronization only on PDSCH DMRS, there is no degradation in performance.

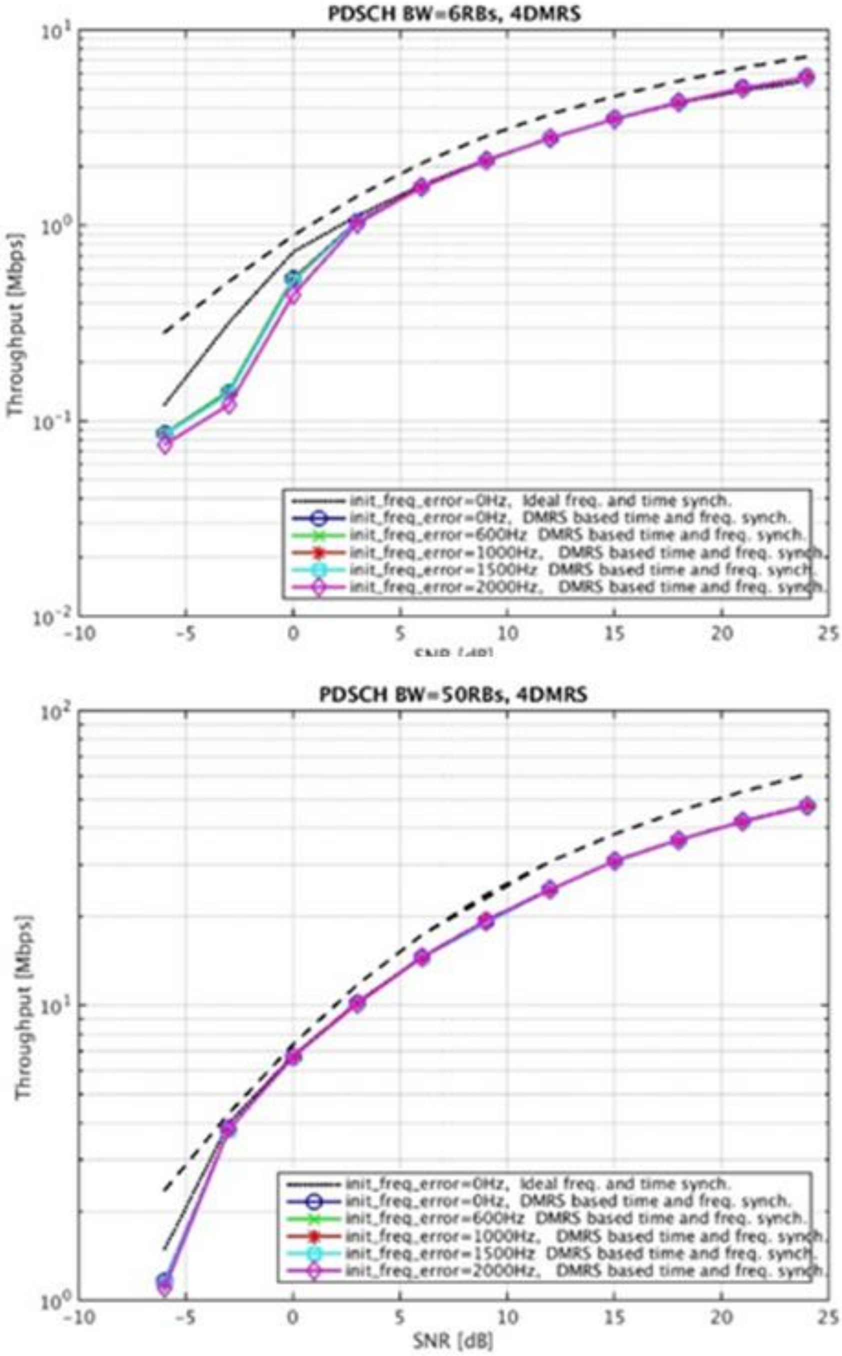

Figure 4 Performance of DMRS only synchronization

Also, the performance of the 2-DMRS per slot PDSCH with 24 RBs allocation of PDSCH performs very similar TRS based synchronization. One reason why it is so good is that the DMRS density is twice the TRS density.

Observation 2 For wide scheduling PDSCH bandwidths, evaluations shows that there is no benefit for UE to be QCL with a TRS, since it can perform fine synchronization using PDSCH DMRS only

This fact indicates a potential for a significant simplification of multi-TRP deployment and operation, especially in FR1. It means that true self-contained PDSCH reception can be used, where the UE performs fine sync and demodulation using the DMRS only. From the network side, the TRP to be used for PDSCH transmission is transparent to the UE and there is no need to indicate the associated TRS that is transmitted from that TRP. This means TRP switching and thus load balancing, interference management can be used, without invoking the QCL framework. We make the following notes:

- We still believe TRS and TCI states is needed for smaller bandwidths, very low SNR reception, so this "TRS-less" reception is for the data offloading using high PDSCH bandwidths. It is for the large packet delivery this load balancing and interference management by fast TRP switching is most useful anyway.
- Likewise, CORESETs will still have TCI state configured, with a TRS. So at least for common search space PDCCH, this fast and transparent TRP switching is not used.
- Since UE needs to perform fine sync on DMRS, there may be a need to revise processing timeline if the UE has not been scheduled for a while and thus need to buffer and adjust the CFO before demodulating the slot.

Proposal 9 Specify support for a wide bandwidth scheduled PDSCH without a QCL to TRS to significantly simplify multi-TRP operation with dynamic point selection.

2.5 Mobility enhancement evolution

In Rel-17 FeMIMO, one of the objectives is to specify L1/L2-centric inter-cell mobility. The underlying idea is to apply the beam management procedures also between TRPs broadcasting different PCIs. This would enable highly efficient UE movement over areas that are not limited to one cell. The progress on this topic is so far slow despite its obvious performance benefits.

Still, even if the work in Rel-17 is completed according to plan, only the initial functionality will be specified, and there are important enhancements that should be considered for Rel-18:

- **Inter-DU operation:** The Rel-17 functionality will be limited to intra-DU operation. This limits the area over which the UE can move without interrupts. Overcoming this limitation would significantly increase the usefulness of the functionality.
- **Enhanced or streamlined RRC signalling:** The Rel-17 work is very much RAN1-centric and will include mainly enhancements to TCI states and L1 reporting. The possibilities to maintain a consistent RRC configuration over larger areas may be addressed during Rel-17, but it is unlikely that it will be optimized. Improving or streamlining the RRC configuration procedures when L1/L2-centric mobility is used over large areas would be an interesting topic to pursue for Rel-18.
- **More flexible operation:** The Rel-17 work focuses on reusing the beam management procedures for inter-PCI operation, and only apply small changes to the framework. In particular, the Rel-17 work focuses on extending the RRC configurations for TCI state and CSI reporting settings. Although straightforward, this will require that many parameters are preconfigured, requiring interaction with RRC just to maintain up-to-date measurement configurations. It would be preferable to maintain the QCL relations and measurement configurations without RRC involvement.
- **Richer reporting:** A beam report can only contain a small number of measurements, and the contents are limited to what is configured by the NW. In many cases, it would be better if the UE could report more RSRP values, and also report neighbours that are not pre-configured by the NW. To limit the UL load, event-driven reporting can also be considered.

Summing up, we propose

Proposal 10 Investigate, and if necessary, specify enhancement to L1/L2-centric inter-cell mobility to increase the areas over which the functionality is appliable.

Proposal 11 Investigate and if necessary, specify enhancements to L1/L2-centric mobility that increases the flexibility and reduces the need for RRC pre-configuration of measurement targets.

5 Conclusion

For the Rel.18 MIMO WI, we propose the following:

Proposal 1 **Specify faster than RRC switching of UL waveform.**

Proposal 2 **Specify multi-layer PUSCH for DFT-S-OFDM.**

Proposal 3 **Specify DMRS capacity enhancements for Type 1 DMRS to at least double the number of orthogonal DMRS ports in one OFDM symbol.**

Proposal 4 **Study, and if beneficial, specify post-decoding CSI feedback for PDSCH (pending outcome of Rel-17 IIoT/URLLC discussion on 'delta-CQI/MCS').**

Proposal 5 **Specify early CSI reporting in Msg3 triggered by the already existing CSI request bit in the RAR grant, both for contention-based and non-contention-based random access.**

Proposal 6 **Specify mixed mode single-DCI and multi-DCI multi-TRP to support high reliability with improved spectral efficiency for intra-UE multiplexing.**

Proposal 7 **Specify Multi-TRP CSI enhancements for multi-TRP URLLC schemes.**

Proposal 8 **Specify flexible URLLC CSI triggering even when a 'High Latency' CSI is being processed.**

Proposal 9 **Specify support for a wide bandwidth scheduled PDSCH without a QCL to TRS to significantly simplify multi-TRP operation with dynamic point selection.**

Proposal 10 **Investigate, and if necessary, specify enhancement to L1/L2-centric inter-cell mobility to increase the areas over which the functionality is appliable.**

Proposal 11 **Investigate and if necessary, specify enhancements to L1/L2-centric mobility that increases the flexibility and reduces the need for RRC pre-configuration of measurement targets.**

5 References

[1] RP-211257, "On CSI Feedback Enhancements for Enhanced URLLC/IIoT", Ericsson, 3GPP TSG-RAN Meeting#92e, June14 – 18, 2021.