3GPP TSG RAN WG1 #90bis **R1-1718580**
Prague, Czech Republic, October 9th – 13th, 2017

Agenda item: 7.3.4.1

Source: Qualcomm Incorporated

Title: Open Issues on BWP

Document for: Discussion/Decision

1 Background

In RAN1 #86bis meeting, bandwidth adaptation facilitating DL control information monitoring over a narrower bandwidth was agreed. The motivation was primarily driven by UE power saving. In the same meeting, support for cross-slot scheduling for DL was also agreed.[1]

The term "bandwidth-part" (BWP) was introduced to refer to the portion of the bandwidth within which the UE operates.

In RAN1 NR AH#3, there was an offline discussion followed by email discussions on the remaining issues on BWP.

In this contribution, we will review some aspects of UE power saving and the connections to BWP, and discuss some of the remaining issues on BWP.

2 Discussion

2.1 UE Power Saving in Active State

One of the main motivations for bandwidth part is UE power saving. It is primarily focused on adaptation of frequency domain bandwidth to optimize for power consumption. As the configuration and signalling framework for BWP is taking shape, support for time-domain power saving techniques should also be considered. From UE power saving and implementation perspective, frequency-domain and time-domain techniques should be jointly considered, because the voltage, clock frequency, and hardware throughput design are often intertwined. If these can be integrated, NR would have a common framework for UE power saving, which would significantly decrease the risk that the devices built to support the first release of NR would be inferior in terms of UE power efficiency compared to LTE.

2.1.1 PDCCH Monitoring Periodicity and Scheduling Parameter k0

PDCCH Monitoring Periodicity

Monitoring PDCCH in every slot facilitates the smallest scheduling latency. Generally, there is a tradeoff between latency and power consumption. With longer time interval between PDCCH occasions, in addition to the direct gain from duty cycle reduction, UE implementation may be able to go into deeper sleep in between. The PDCCH monitoring periodicity should be configurable to a multiple of slots to allow this tradeoff. Especially for very short slot duration (e.g. for 120kHz SCS, slot duration is only 0.125 milliseconds), gNB and UE can agree that PDCCH should be monitored only at certain slot periodicity, even during active state, in order to save power. One drawback is that some slots may not be schedulable.

Observation 1: Larger PDCCH monitoring periodicity reduces power consumption due to reduced duty cycle.

RAN1 has already agreed that at least one CORESET is configured per BWP [5]. PDCCH monitoring periodicity is an attribute of CORESET configuration. Therefore it is already implied that PDCCH monitoring periodicity configuration can also be BWP-specific, and it can be configured as part of CORESET configuration associated with a BWP.

Microsleep

Microsleep is a technique to put some of the modem's hardware into sleep mode for the data portion of a slot (or subframe in LTE), if no DL grant has been decoded for the slot. However, the extent of power saving that can be

achieved with microsleep for LTE is fairly limited. First, due to short time scale of the sleep, only a limited number of hardware blocks (e.g. RF circuitry) can be put into sleep mode. Second, there is dependency on the same slot grant, which is not decoded until at least several symbols after the last DL control symbols. For NR, DL control channel decoding complexity is likely to be improved from LTE; However, given that the main use case for subcarrier spacing is likely to be 30kHz or higher, the slot duration is also going to be reduced to half of the subframe duration or even less. This poses challenges for the extent of power saving that microsleep can achieve.

Cross-Slot Scheduling

If the grant can be transmitted at least one slot in advance, i.e. data assignment being cross-slot scheduled, DL control channel processing would not be in the critical timeline for microsleep decision. Suppose in Slot n, UE knows from decoding the PDCCH that there is no grant for the Slot n+1; During Slot n+1, it only needs to buffer up first few symbols containing the PDCCH, and can immediately go into microsleep for the rest of the slot. During microsleep in Slot n+1, the modem can still operate in low power mode and process the PDCCH based on the captured Rx samples. Overall, the duration achievable for microsleep would be close to the entire data portion of the slot. Similar observations have been confirmed in another company's contributions [2][3].

Cross-slot scheduling, especially across multiple slots, and configuration of sparser PDCCH monitoring periodicity tend to increase scheduling latency. On the other hand, such techniques tend to allow further power consumption reduction. In addition to better microsleep, cross-slot scheduling opens up opportunities for additional power saving if UE implementation takes advantage of it.

To understand why additional power saving can be attained with cross-slot/cross-multiple-slot scheduling, it is important to be aware of an VLSI technique called DVFS (dynamic voltage frequency scaling). With DVFS, the voltage and clock frequency of a digital circuit can be dynamically adapted to fit the tasks at hand, and significant power saving ensues compared to always running the voltage and clock frequency at the peak settings. However, there is a physical constraint for how fast the adaptation can happen.

In a grossly simplified view, the modem could be considered the "circuit" under DVFS operation. The workload for different modem tasks, such as PDCCH monitoring, and/or PDSCH reception, can be very different. As a result, the ideal operating point for DVFS can be different corresponding to different modem tasks. However, the modem implementation may not support instantaneous adaptation of DVFS operating point that is perfectly aligned to the modem task at hand. The modem may have to operate at a "worst case" DVFS operating point (i.e. that consumes more power than ideal) in anticipation of the possible modem tasks that may come within the adaptation time scale.

As an example, with same-slot scheduling (k0=0), UE may be getting a grant during PDCCH monitoring at the beginning of a slot, and would need to be prepared to start processing PDSCH in the same slot. The DVFS operating point may need to be set to the worst case of decoding a grant and processing PDSCH, which would be higher than simply monitoring PDCCH alone.

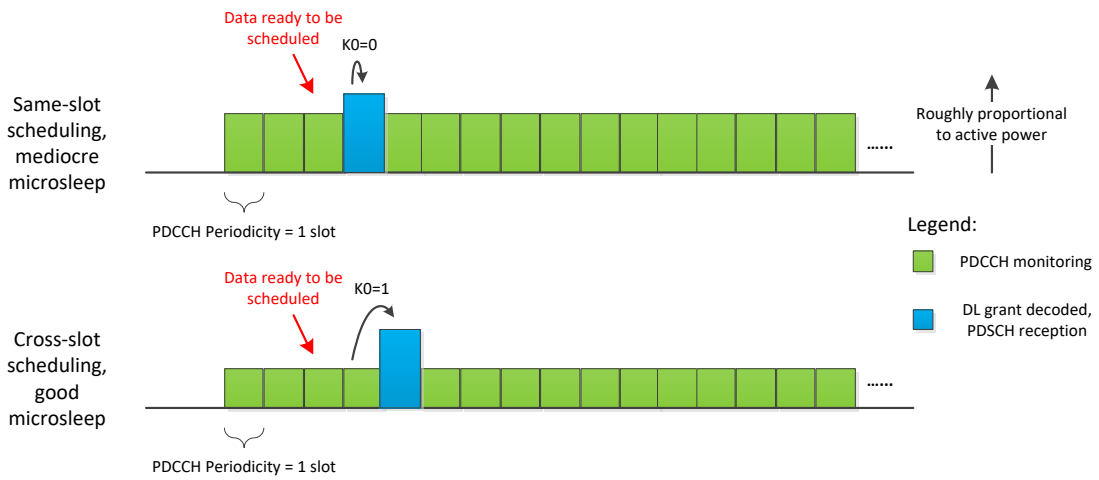

Figure 1. Scheduling latency vs. power consumption tradeoff (1): Cross-slot scheduling facilitating better microsleep and DVFS

With cross- slot scheduling with k0 larger than 1, additional power saving could be larger due to the extent of DVFS that becomes feasible. Moreover, if the modem tasks during certain time duration can be limited, modem architecture

can be designed in such a way that the limited functionality could operate on a highly optimized block. For example, if k0 is sufficiently large, a highly power-optimized PDCCH processing block can be used to further reduce PDCCH monitoring power consumption, while the rest of the modem can remain in idle/very low power mode. When a DL grant is decoded, the modem still has enough time to "warm-up" and get ready for PDSCH reception and HARQ-ACK feedback.

If k0 can be dynamically selected via DCI, it is important that the set of selectable values, which are semi-statically configured, should not be smaller than a threshold required for modem "warm-up". UE implementation can take advantage of the fact that even the smallest k0 which is dynamically selected would be large enough for modem "warm-up" for full functionality, so that PDCCH monitoring can operate with limited functionality and at the lowest power level possible.

Observation 2: Sufficiently large k0 value allows UE modem to monitor for PDCCH with highly optimized mode or hardware, and potentially operate at fraction of the power level compared to the k0=0 case.

While it has already been agreed that k0 can be at least semi-statically configured for a UE, it makes sense to discuss whether and how this agreement can be extended for BWP.

Proposal 1: Semi-static configuration of k0 should be BWP-specific, and it can be configured as part of BWP configuration.

The motivation and benefits are clear: Given that k0>=1 is good for power saving, it makes sense to configure that for the BWP which is intended for low power operation. For example, in the case with a narrower BWP for DL control monitoring and a wider BWP for scheduled data reception, cross-slot scheduling (i.e. k0>=1) could be configured for the narrower BWP, and same-slot scheduling (i.e. k0=0) could be configured for the wider BWP.

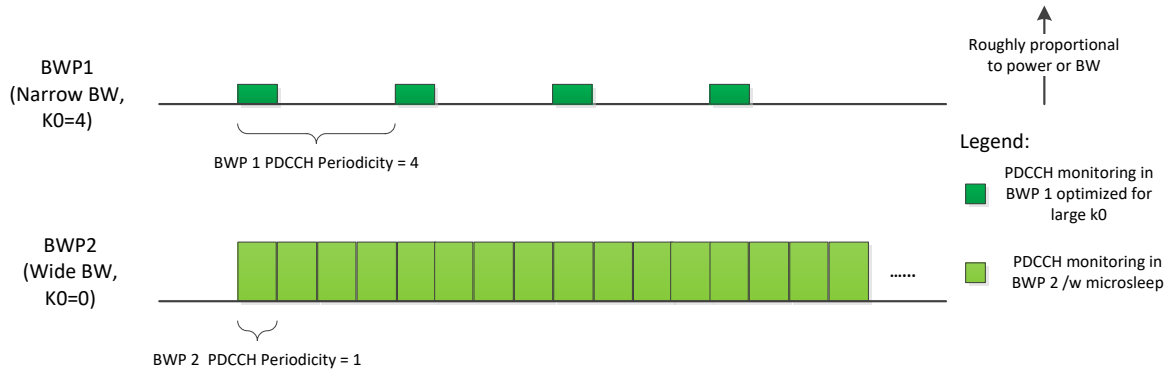

Figure 2. An example of BWP configuration with PDCCH monitoring periodicity and k0

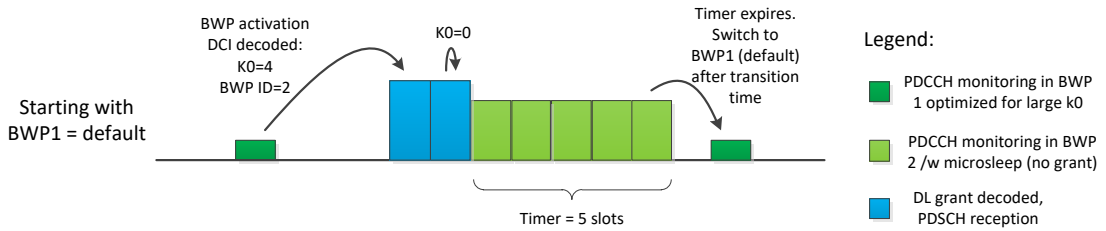

Figure 3. An example of UE monitoring PDCCH in BWP1 and switching to BWP2 due to a scheduling DCI

Figure 2 illustrates k0 and PDCCH periodicity are both set to the same value of 4 for BWP1. In general, there are two other cases to consider: (i) k0 > PDCCH periodicity, (ii) k0 < PDCCH periodicity. For (i), scheduling overlap may result, but it is still a valid mode of operation. For (ii), there could be extra "wake-ups" of the UE to receive PDSCH during sleep mode in between PDCCH occasions.

It should be understood that the configuration of PDCCH monitoring periodicity is also associated with an absolute definition in time for the PDCCH occasions with respect to some common reference timing, so that gNB and UE would not be misaligned in terms of the PDCCH occasions. This is crucial for robustness considerations.

2.2 Remaining Issues on BWP

2.2.1 Relationship between CA and BWP

The relationship between CA and BWP pertains to issues such as the configuration of cell and how BWP can be configured and associated with the cell. For SCell, the issue of SCell activation/deactivation and how it relates to BWP activation/deactivation has to be resolved. These issues are discussed in our companion contribution [9].

2.2.2 BWP Activation/Deactivation without Scheduling

In RAN1 NR AH#3, the following agreement was made [8]:

- NR supports the case that a single scheduling DCI can switch the UE's active BWP from one to another (of the same link direction) within a given serving cell
    - FFS whether & how for active BWP switching only without scheduling (including the case of UL scheduling without UL-SCH)

The following discussion will motivate the need and use cases for dedicated BWP DCI for active BWP switching without scheduling.

Support for a single scheduling DCI to trigger active BWP switching is motivated mainly by dynamic BWP adaptation for UE power saving during active state (which includes ON duration and when inactivity timer is running when C-DRX is configured). Based on DoU profiling results, even with C-DRX enabled, UE typically consumes significant amount of power monitoring PDCCH without decoding any grant. [1]. To reduce the power consumption during PDCCH monitoring, the main use case is to have two BWPs configured: a narrower BWP mainly for PDCCH monitoring, and a wider BWP mainly for scheduled data. In such use case, it is envisioned that the UE may switch back-and-forth between the narrower BWP and the wider BWP, depending on the burstiness of the traffic. Most of the time, UE would be "revisiting" a BWP that it has dwelled on previously. For this case, we believe there is merit to allow BWP switching indication and scheduling grant to be combined for low latency and reduced signalling overhead for BWP switching. The performance loss for not having the latest CQI immediately following the switch may be manageable most of the time.

There are other use cases where a dedicated BWP DCI for switching/activation/deactivation would be beneficial. If UE does not have prior CQI information on a BWP (or it has been very stale), it would be more efficient for UE to switch to the new BWP, perform CSI measurement and feedback, and then gNB can schedule to the UE based on an up-to-date CQI. This scheme avoids the complexity for supporting CSI measurement outside of configured BWP. While BWP switching triggered with a scheduling grant could be the main use case, support for above use case is also very important to ensure robust operation of the system.

Another use case is for BWP activation/deactivation for SCell. It has been agreed that for Rel 15, only single active BWP (for each link direction) is supported. Support for separate BWP activation and deactivation signaling, at least for PCell, does not seem necessary because PCell has to always maintain an active BWP for each link direction; Therefore BWP switching is more applicable. However, for SCell, it is not fully clear that separate BWP activation and deactivation signalling is not needed. It is still being discussed whether SCell activation and deactivation triggers the corresponding action for its configured BWP, or vice versa. This topic is discussed more in detail in a companion contribution [9].

According to our understanding, the main concern for supporting a dedicated BWP activation/deactivation DCI is the impact to DCI format. To alleviate the concern, reuse of a scheduling DCI with dummy grant could be considered. A dummy grant may be constructed by invalidating one or some of the fields, such as the resource allocation field. Additionally, it may also be feasible to leverage a fallback scheduling DCI format (which contains a smaller payload) to improve the robustness for BWP DCI signalling, without incurring extra work on introducing a new DCI format.

Proposal 2: Support dedicated BWP DCI for switching/activation/deactivation without scheduling

Proposal 3: Reuse existing DCI formats for the dedicated BWP DCI. Consider reusing the format of scheduling DCI with dummy grant.

2.2.3 Timer for Switching to Default BWP

In RAN1 #90, the following agreement was made [6]:

- Support activation/deactivation of DL bandwidth part by means of timer for a UE to switch its active DL bandwidth part to a default DL bandwidth part
    - The default DL bandwidth part can be the initial active DL bandwidth part defined above
    - FFS: The default DL bandwidth part can be reconfigured by the network
    - FFS: detailed mechanism of timer-based solution (e.g. introducing a new timer or reusing DRX timer)
    - FFS: other conditions to switch to default DL bandwidth part

Based on offline discussion in the AH#3 meeting, there is a slight majority of companies supporting introduction of a new timer, but consensus was not reached. In the following, the merits for a new timer are discussed.

First, we want to point out that reusing a DRX timer for the BWP feature may introduce unnecessary intertwining between RAN1 and RAN2 standardization efforts. DRX timers are specified in RAN2, but BWP is mainly specified in RAN1.

Second, there is a technical reason for not reusing the DRX inactivity timer for BWP switching to the default, which is motivated by UE power saving. For initial deployment of Rel 15, it is envisioned that BWP usage would be very simple: default BWP could be configured to be a narrower BWP intended for PDCCH monitoring at low power consumption, and another wider BWP configured for scheduled data. To minimize signalling overhead, a single scheduling grant can switch the active BWP to the wider BWP from the narrower BWP. After the traffic burst tapers off, it is envisioned that return to the narrower (default) BWP is triggered by the timer without additional signalling. If the same DRX inactivity timer is used, this means UE would stay in the wider BWP for as long as the inactivity timer is running, which could be a long time. Typically, DRX inactivity timer is set to a large value of 100~200 milliseconds for C-DRX cycle of 320 milliseconds, much larger than the ON duration (typically only 10 milliseconds). This implies that for most of the cases, power saving due to narrower BWP is not realized – It is exercised only for the ON duration for which no data is scheduled, or some portion of the ON duration before data is scheduled.

To realize fuller potential of UE power saving promised by BWP switching, a new timer has to be defined and it should be configured to be smaller than the DRX inactivity timer. From the point of view of DRX operation, BWP switching allows UE to operate at different power levels during the active state, effectively providing some more intermediate operating points between the ON and OFF states.

Proposal 4: For timer-based mechanism to switch from an active BWP to the default BWP, a new timer is introduced.

The triggering condition should be based on grant, same as DRX inactivity timer. This is based on the assumption that scheduling grants tend to cluster close together in time; The same principle is assumed for DRX inactivity timer.

RAN2 has decided to base the unit of DRX timers on absolute time (i.e. milliseconds) [7] instead of in units of slots, symbols, or PDCCH monitoring periodicity, which are all numerology dependent. For the new timer, a similar approach can be taken.

Proposal 5: For the triggering condition and definition of the time units for the new timer, adopt the same as the DRX inactivity timer.

2.2.4 Maximum Number of BWP Configuration per Cell

During offline discussion in RAN1 NR AH#3, there has been a near consensus that the maximum number of BWP configuration per cell (and for each link direction) should be 4. From our view, this is mainly based on the accounting that there has to be at least a default BWP, another configured BWP with wider bandwidth, and one or two more for flexibility.

Considering the complexity of BWP and the amount of standardization work, RAN1 should consider introducing the feature with a phased approach.

Proposal 6: For Rel-15 Phase 1, support up to 2 BWP configurations per cell for each link direction.

Allowing up to two configured BWPs should cover the most important use case for BWP switching, which is to dynamically switch between a narrow (default) BWP and a wider BWP. In our view, it is important to prioritize getting

this use case to work. If BWP complexity is too high, there is a real risk that either standardization work cannot complete in time and/or implementation/testing becomes unmanageable for timely deployment.

2.2.5 Common Search Space Support

In RAN1 NR AH#2, the following agreement was made [5]:

- At least one of configured DL BWPs includes one CORESET with common search space at least in primary component carrier
- Each configured DL BWP includes at least one CORESET with UE-specific search space for the case of single active BWP at a given time
    - In case of single active BWP at a given time, if active DL BWP does not include common search space, then UE is not required to monitor the common search space

UE needs to monitor for RMSI and broadcast OSI which is transmitted by the gNB within the common search space (CSS) on the PCell. In addition, RACH response and paging control monitoring on the PCell can also be transmitted within the CSS. Above agreement allows UE to be on an active BWP configured with UE-specific search space (USSS) only, and when that is the case, UE is not required to monitor the common search space.

For a PCell, at least one of configured DL bandwidth parts includes one CORESET with CSS type. It has been proposed that to monitor RMSI and broadcast OSI, UE can periodically switch to the BWP containing the CSS. Likewise for RACH response and paging control monitoring on the PCell.

While this proposal may be workable, if BWP switching to monitor the CSS happens frequently, it introduces overhead and compromises the intended UE power saving objective.

Between any two BWP, there are three scenarios regarding overlapping in frequency, as illustrated below:

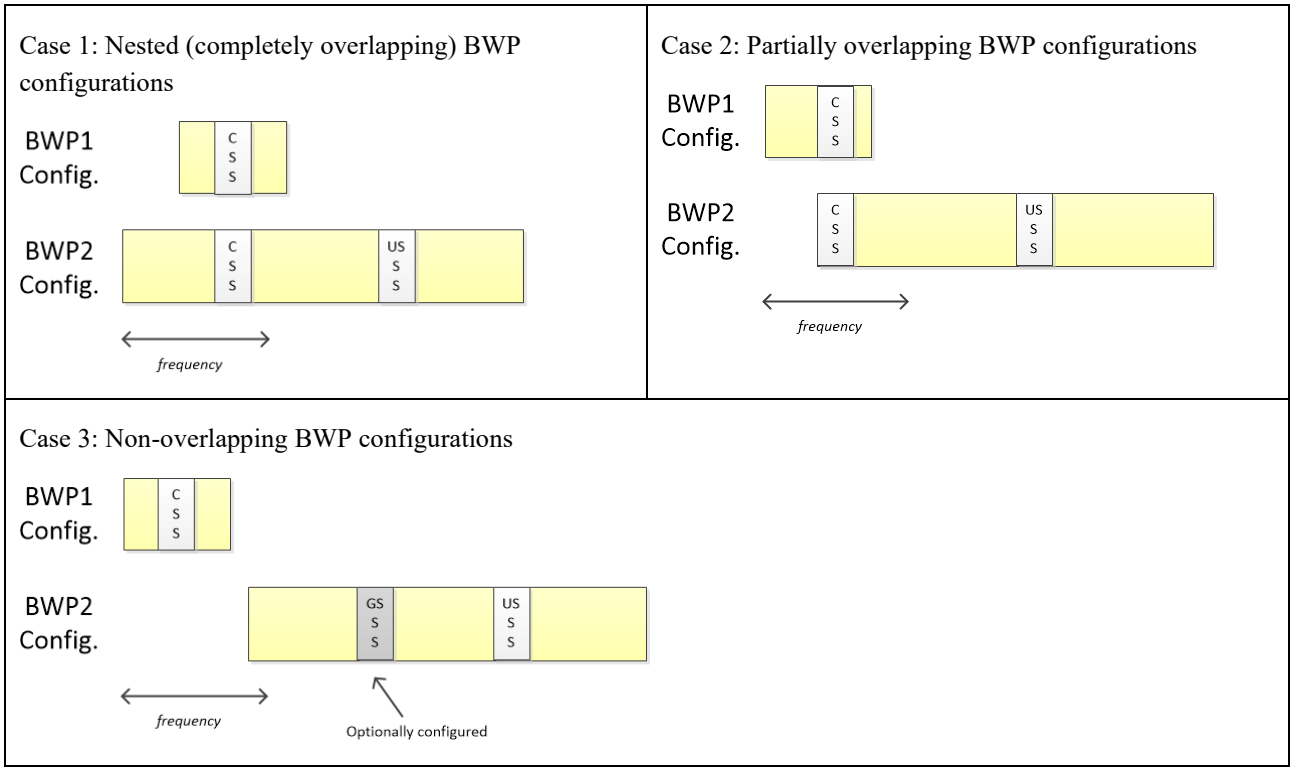

Figure 4: Scenarios for BWP configurations overlapping in frequencies

For Case 1 of nested BWP configurations, the same CORESET configuration can be used across the BWPs. Unless reconfigured otherwise, the default BWP is also the one containing the CSS, as illustrated as BWP1 in Case 1 of Figure 4. If follows that BWP2 also contains the CSS. In Case 2 of Figure 4, the BWPs are partially overlapping. If the

overlapping region is sufficient, BWP2 can be configured to include the CSS; Otherwise, from the perspective of configuring the same CORESET containing the CSS, it can be treated as non-overlapping (Case 3).

There are multiple benefits of configuring the same CORESET containing the CSS across BWPs. First, RMSI and broadcast OSI monitoring can be handled without necessitating BWP switching. Likewise, RACH response and paging control monitoring on the PCell can also be handled without switching. Second, if CORESET configuration is the same across BWPs, robustness for BWP switching would improve, because even if gNB and UE are out-of-sync as to which BWP is currently active, the DL control channel still works (Note: This requires a further requirement that at least some DCI format should be functionally invariant across BWP configurations. Functional invariance means the format size is the same; Some fields can have different format as long as the UE can interpret them based on the content in other fields which have the same format). The constraints on BWP configuration may not be too much, considering that BWP is intended for power saving, even the nested configuration should be very versatile for different applications.

Proposal 7: For Rel-15 Phase 1, for PCell, only support the BWP configurations for a UE for which all BWP can be configured the same CORESET containing the CSS.

- **Broadcast information including at least RMSI and OSI can be received by the UE without necessitating switching to a particular BWP**
- **At least one of DCI formats should be functionally invariant across BWP configurations**

For the scenarios where the BWP configurations are non-overlapping in frequency (Case 3), extending the aforementioned agreement, there should not be spec mandate for UE to monitor RMSI and broadcast OSI in the CSS. It is left to implementation to handle this case.

NR also supports group-common search space (GCSS) and this can be used as an alternative to CSS for certain information. gNB can configure GCSS within a BWP for a UE, and information such as RACH response and paging control can be transmitted on GCSS, and UE can monitor GCSS instead of switching to the BWP containing the CSS for such information.

For pre-emption indication and other group-based commands on a serving cell, similar to above, gNB can transmit the information on GCSS. UE may monitor the GCSS for the information. This is especially useful for SCell which may not have CSS.

Proposal 8: Group-common search space can be configured within a BWP for transmission of certain types of broadcast information including RACH response and paging control, and/or group-based information including pre-emption indication.

3 Conclusions

We discussed some aspects of UE power saving and the connections to BWP, and some of the remaining issues on BWP. The following observation and proposals have been made:

Observation 1: Larger PDCCH monitoring periodicity reduces power consumption due to reduced duty cycle.

Observation 2: Sufficiently large k0 value allows UE modem to monitor for PDCCH with highly optimized mode or hardware, and potentially operate at fraction of the power level compared to the k0=0 case.

Proposal 1: Semi-static configuration of k0 should be BWP-specific, and it can be configured as part of BWP configuration.

Proposal 2: Support dedicated BWP DCI for switching/activation/deactivation without scheduling

Proposal 3: Reuse existing DCI formats for the dedicated BWP DCI. Consider reusing the format of scheduling DCI with dummy grant.

Proposal 4: For timer-based mechanism to switch from an active BWP to the default BWP, a new timer is introduced.

Proposal 5: For the triggering condition and definition of the time units for the new timer, adopt the same as the DRX inactivity timer.

Proposal 6: For Rel-15 Phase 1, support up to 2 BWP configurations per cell for each link direction.

Proposal 7: For Rel-15 Phase 1, for PCell, only support the BWP configurations for a UE for which all BWP can be configured the same CORESET containing the CSS.

Proposal 8: Group-common search space can be configured within a BWP for transmission of certain types of broadcast information including RACH response and paging control, and/or group-based information including pre-emption indication.

4 References

[1] R1-1610135, "Evaluation of Frame Structure Design for UE Power", RAN1#86bis, Qualcomm
[2] R1-1609556, "Cross-slot scheduling in NR", RAN1#86bis, Mediatek
[3] R1-1609557, "On Techniques for Enhanced UE Power Efficiency", RAN1#86bis, Mediatek
[4] Chairman's Notes 3GPP TSG RAN WG1 #86bis
[5] Chairman's Notes 3GPP TSG RAN WG1 NR Ad-Hoc #2, 2017
[6] Chairman's Notes 3GPP TSG RAN WG1 #90
[7] User Plane Meeting Minutes, 3GPP TSG RAN WG2 #99
[8] Chairman's Notes 3GPP TSG RAN WG1 NR Ad-Hoc #3, 2017
[9] R1-1718581, "Open Issues on CA", RAN1#90bis, Qualcomm