IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received May 8, 2020, accepted May 30, 2020, date of publication June 9, 2020, date of current version June 22, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3001180

Power Saving Techniques for 5G and Beyond

YU-NGOK RUYUE LI[1], (Member, IEEE), MENGZHU CHEN[1], JUN XU[1], LI TIAN[1], (Member, IEEE), AND KAIBIN HUANG[2], (Senior Member, IEEE)

[1]ZTE Corporation, Shenzhen 518057, China
[2]Department of Electrical and Electronics Engineering, The University of Hong Kong, Hong Kong

Corresponding author: Yu-Ngok Ruyue Li (li.ruyue@zte.com.cn)

ABSTRACT Energy efficiency is one of the key performance indicators in 5G New Radio (NR) networks targeted to support diversified use cases including enhanced mobile broadband (eMBB), massive machine type communications (mMTC) and ultra-reliable and low latency communications (URLLC). Trade-offs have to be carefully considered between energy efficiency and other performance aspects such as latency, throughput, connection densities and reliability. Energy efficiency is important for both user equipment (UE) side and base station side. On UE side, UE battery life has great impact on user experience. It is challenging to improve UE experience in other performance aspects without affecting battery life of 5G handsets. On the base station side, efficient network implementation is critical in both environmental and operation cost standpoints. To adapt different requirements and trade-offs, the 5G NR standard is designed to have great flexibility on network operation modes. This paper provides an overview on power saving techniques supported by 5G NR standards according to the current 5G standardization progress. It provides the 5G evolution path of the power saving techniques from the first release of 5G standard to the future beyond-5G releases. In addition to the existing standardized techniques, some major development trends of green communication and the future potential enhancements expected in the beyond-5G standards are discussed.

INDEX TERMS Power saving, beyond 5G, energy efficient network, green communication.

I. INTRODUCTION

The Third Generation Partnership Project (3GPP) completed the 5G standardization process of Release 15 New Radio (NR) access technology which is the first version of the 5th Generation wireless standard [1]. Meanwhile, the standardization process of next 5G release i.e. Release 16 is still on-going and will be expected to finish in June 2020. The plan for the third 5G release i.e. Release 17 has been discussed and topics to be worked on have been identified in the plan [2]. Compared to 4G Long Term Evolution (LTE), 5G NR standard aims to support more diversified use cases. Major 5G use cases can be generalized to three usage scenarios, namely enhanced mobile broadband (eMBB), massive machine type communications (mMTC) and ultra-reliable and low latency communications (URLLC) [3]. For different use cases, there are corresponding key performance indicators. For example, user experienced data rate is important for eMBB. Latency and reliability are the key parameters for URLLC traffic. While for mMTC, it focuses more on connection density. One important aspect for all traffic types is energy efficiency which is a key performance indicator (KPI) on both UE side and network side.

The impact of energy saving techniques on network energy efficiency, which is defined by bits per Joule, is analyzed in [4], [5]. Solutions including multiple input multiple out (MIMO), heterogeneous networks (HetNets), non-orthogonal aggregation (NOA), and exploiting unused and unlicensed spectrum, etc., are provided to improve system throughput. RF energy reduction is proposed to achieve a trade-off between energy efficiency and throughput [5]. In some scenarios, e.g. massive MIMO, power consumption is still expected to be one of the bottlenecks on both UE and network sides in early 5G deployment Hence, standardization support of enhancements on power saving techniques are essential to realize these techniques in practical network deployment of 5G and beyond.

Network energy efficiency for different deployment scenarios and traffic characteristics has been taken into account in NR design [3], [6]. In NR Release 15, an energy efficient 5G standardized framework has been built with high flexibility and scalability. Several features make the system more

The associate editor coordinating the review of this manuscript and approving it for publication was Ilsun You.

flexible to adapt to different traffic loads for better energy efficiency. In 4G LTE, always-on CRS (cell specific reference signal) for channel estimation impose restrictions on energy efficient network implementation and generates unnecessary interference. Some kind of small cell muting schemes were proposed to reduce interference and power consumption in LTE [7] but it is hard to address this issue in LTE considering the backward compatibility constraint. In 5G NR standard [8], redesign of reference signal is possible. The flexibility and configurability of reference signal design in 5G minimizes always-on signals. In addition, multiple options of rate-matching resources, e.g. resource block (RB) level or resource-element (RE) level, dynamic or semi-static, enable flexible muting of resources and ensure forward compatibility for energy efficient network implementation.

With the flexibility provided by the standardized framework, the trade-off between power saving and performance including throughput and latency can be more easily controlled by network implementation. Regarding power saving on UE side, some power saving schemes like discontinuous reception (DRX) [9] mechanism are inherited from 4G LTE. However, it is not sufficient to only reuse the schemes from 4G LTE due to more diversified use cases and scenarios in 5G NR. Hence, the standardized framework in NR provides more flexibility to allow the network to implement different kinds of adaptation based on the power saving need from UE side and network side. These different kinds of adaptation includes adaptation to bandwidth, MIMO layers, control channel monitoring, etc. Brief overview on the standard evolution of different techniques, including UE power saving, towards 5G-advanced is provided in [10] but it lacks the details for providing a more complete picture on standardization support of power saving techniques.

This paper provides an overview of 5G evolution path on recently standardized power saving techniques in more details. In addition to flexible reference signal design, some essential standardized components for energy efficient 5G network supported in Release 15 NR standard are described in Section II, including bandwidth adaptation, Radio Resource Control (RRC) inactive state, DRX mechanism, control channel design and cross-slot scheduling. In Section III, details of enhancements on power saving introduced in Release 16 are provided. Simulation results showing power saving gain of these enhancements are provided. Future trends of power saving enhancements considering beyond 5G are discussed in Section IV.

II. STANDARDIZED FRAMEWORK FOR ENERGY EFFICIENT 5G SYSTEM

The energy efficiency can be expressed as the following formula without considering different deployment scenarios [3], [11]

$$EE = \sum_i \alpha_i \frac{V_i}{E_i} \tag{1}$$

where for each traffic load level i V_i denotes the traffic load per second processed by a wireless device, E_i refers to the power consumed by the wireless device to process the traffic load, α_i is the weight of traffic load level i. The definition of energy efficiency can be applied to both base station and user equipment (UE). According to the formula, the energy efficiency can be enhanced by boosting up the traffic load using techniques, such as utilizing wider spectral resources, more antennas, higher modulation order, provided that power consumption to process the increased traffic load can be kept low. Therefore, energy saving is a key factor for sustainable energy-efficient 5G network deployment. This paper focuses on improving energy efficiency by reducing network and UE power consumption. In addition to UE implementation, UE power consumption relies on the configurations by network, which needs to be guaranteed by specification. Techniques that are aimed to reduce UE power consumption are mainly analyzed in this paper.

In this section, some essential components for building a standardized framework of energy efficient 5G system are provided. These components include flexible reference signal design, efficient sleep modes, bandwidth adaptation, RRC inactive state, DRX mechanism, control channel design and cross-slot scheduling. Some of the description is more from UE power saving perspective. Some techniques can be used for power saving from network perspective in combination with flexible and scalable framework as described in Section I.

A. FLEXIBLE REFERENCE SIGNAL DESIGN FOR EFFICIENT SLEEP MODE

As mentioned in Section I, always-on reference signals on LTE has limited the energy efficiency of 4G network. In contrast, flexible reference signal design in 5G enables more efficient sleep mode at the base station. Without always-on reference signals, the base station can just wake up periodically to send synchronization signals, broadcast necessary system information and detect random access channel (RACH). Here we briefly discuss on two mandatory downlink signals and the possible periodicities in 5G NR [8].

- *Synchronization Signal Block (SSB):* Multi-beam based SSB is defined in 5G NR to include Primary SS (PSS), Secondary SS (SSS) and Physical Broadcast Channel (PBCH) transmitted by the base station periodically. Periodicity of SSB is assumed to be not longer than 20ms for standalone NR cases, i.e. it can be 5ms, 10ms or 20ms. While for non-standalone case with LTE-NR Dual Connectivity, the periodicity can be {5, 10, 20, 40, 80, 160}ms.
- *Channel State Information Reference Signal (CSI-RS):* CSI-RS can be used for channel tracking, CSI acquisition, mobility or beam management. Periodic CSI-RS for channel tracking is mandatory for UEs under connected state to perform fine time and frequency channel tracking. It is transmitted in bursts of two or four symbols which are spanned across one or two consecutive slots. Periodicity of this channel tracking RS can be

flexibly configured as {10, 20, 40, 80}ms. One energy efficient implementation is to configure long periodicity such 80ms for periodic CSI-RS for tracking. Once traffic comes, aperiodic tracking RS can be triggered before data so that the UE can perform fine channel tracking before data reception and demodulation.

The flexible periodicity of reference signal design in NR is a significant improvement on network energy efficiency compared to LTE in which CRS has to be transmitted in every 1ms. With this flexibility, 5G base station can support different levels of sleep mode depending on the traffic load. In [12], four levels of sleep modes in the power model are defined to represent different depths of sleep modes measured by transition latency including de-activation and re-activation latency on the sleep mode. The four levels respectively correspond to OFDM symbol level (i.e. 71.4us), sub-frame level (i.e. 1ms), radio frame level (i.e. 10ms) and the deep sleep stand-by level in the order of 1s. In LTE, OFDM symbol level is the only level it can use due to always-on reference signals. In NR, four levels of sleep modes can be achieved by different periodicity of reference signals and power saving schemes like Secondary Cell (SCell) dormancy for carrier aggregation.

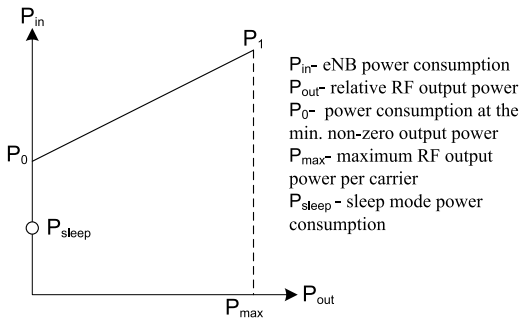

FIGURE 1. Load-dependent base station power model [13].

Energy efficiency can be evaluated using the data from the recent power model in [12] together with the simplified estimate of a power model for base station proposed in [13], [14] as shown in Figure 1 and the formula (2) below. The base station power consumption P_{in} can be obtained from the following function:

$$P_{\text{in}} = \begin{cases} N_{CC} N_{\text{sec}} \cdot (P_0 + \Delta_{\text{p}} P_{\max} \chi) , & 0 < \chi \leq 1 \\ N_{CC} N_{\text{sec}} \cdot P_{\text{sleep}} , & \chi = 0 \end{cases} \quad (2)$$

where N_{CC} is the number of component carriers (CC), N_{sec} is the number of sectors per site, Δ_{p} is the slope of the load dependent power consumption, $\chi = [0, 1]$ is the ratio of the number of transmitted resource elements (REs) to the total number of REs.

Using this model and the updated parameters for 5G base station in Table 1, energy saving performance is evaluated using system level simulation on small cell deployment with different densities. Different cell densities are achieved by deploying different number of small pico cells (N_cell) per

TABLE 1. Power model parameters for pico small cell [12].

BS type	N_{sec}	$P_{\max}$ [W]	$P_{\max}$ [dBm]	P_0 [W]	Δ_{p}	P_{sleep} [W] for sleep modes
Pico	1	1.0	30	2.3	4.6	1.5 (L1) 0.4 (L2) 0.2 (L3)

macro area. To compare power consumption used with different densities and sleep modes, the same traffic load using FTP model 1 [15] with file size = 0.5MByte and packet arrival rate $\lambda = 15$ per second per macro area is used for all simulations. More detailed simulation assumptions can be found in [8] as similar simulation setup is used here. In the simulation, sleep modes are applied to small cells only. With this simulation setup, energy saving of level 2 (L2) and level 3 (L3) sleep modes are obtained as shown in Figure 2, compared to the baseline 4G LTE deployment which level 1 (L1) sleep mode is used due to always-on reference signals.

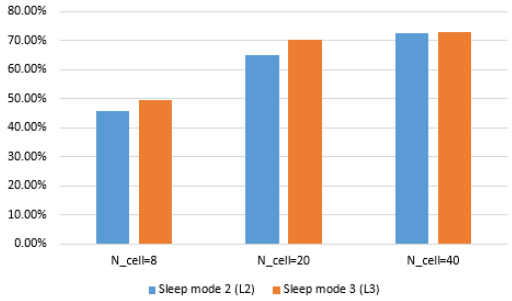

FIGURE 2. Energy saving percentage gain with different sleep modes.

It can be observed from the results that more energy can be saved for denser small cell deployment. The energy saving can be more than 70% for the scenarios where 20 or 40 small cells are deployed per macro area, compared to 4G LTE deployment. With denser small cell deployment, resource utilization is lower which favors energy saving due to more frequent sleep. In 4G LTE, the performance of energy saving is limited by level 1 sleep mode and the extra overhead and interference due to always-on reference signals. Therefore, flexible reference signals in 5G has enabled more energy efficient dense network deployment.

B. BANDWIDTH ADAPTATION

The key performance requirements of 5G technologies by IMT-2020 include downlink peak data rate is 20Gbit/s and uplink peak data rate is 10Gbit/s. To achieve the requirements, a wide bandwidth is supported by the new radio access technique developed by 3GPP. According to 3GPP specification, UE channel bandwidth is up to 100MHz for sub-6GHz bands [16] and 400MHz for above 6GHz bands [17], which is much wider than the bandwidth of 20MHz in LTE. It should be noted that UE capability of supported bandwidth varies and is often limited to the bandwidth less than the maximum

bandwidth supported in the specification especially in the early stage of 5G deployment. Besides, high UE power consumption can be a major issue if UE is required to perform transmission or reception in a wide bandwidth all the time regardless of how much the actual traffic load is.

To reduce UE power consumption and guarantee the data transmission rate, the concept of bandwidth part (BWP) was adopted by 3GPP. A BWP is comprised of a number of continuous physical resource blocks (PRB) with specific numerology. For each serving cell, there are at most four BWPs can be configured for downlink (DL) or uplink (UL). Only one UL BWP and one DL BWP are active at a given time instant. Furthermore, UE is not required to transmit or receive data outside an active BWP. The BWP can be activated or de-activated by a timer, physical layer Downlink Control Information (DCI) signaling or higher layer RRC signaling. When a large data packet needs to be transmitted, UE can be indicated to activate a BWP with a wide bandwidth. Otherwise, UE can be informed to switch to a BWP with a narrow bandwidth to save power [18]. BWP switch delay is defined with regard to the activation signaling. An example is shown in Figure 3.

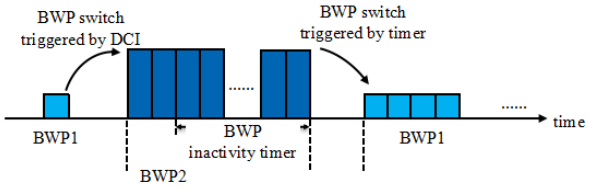

FIGURE 3. BWP adaptation.

Overall, standardized BWP switching framework in NR is beneficial for building an energy efficient network. In addition to bandwidth adaptation, this can also serve as a general framework which can be used for adaptation of other power saving parameters introduced in the 5G evolution.

C. RRC INACTIVE STATE

In LTE, RRC connected state and RRC idle state are defined so that UE can operate under low power mode in idle state when there is no data transfer. When there is data activity, UE transitions back to RRC connected state by going through a cumbersome procedure described in [19] which requires extensive signaling between UE and the network. Inactivity timer is set to trigger UE to idle state if there is no data activity in a certain duration. The timer setting can be considered as a tradeoff between power consumption and data transmission efficiency including signaling overhead and latency. The timer setting with short duration favors more on idle state but it costs higher signaling overhead and latency if there is frequent data transmission. However, the power saving gain from idle state is reduced if the timer expires only after a long duration. The major issue of this state transition is high signaling and latency cost.

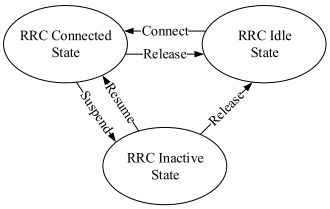

FIGURE 4. NR RRC state machine.

To address this issue, a new state called RRC inactive state [1], [19] has been introduced in NR, in addition to RRC connected and RRC idle states as shown in Figure 4. The motivation of this new RRC state is to allow faster and more efficient resumption to RRC connected state so that data transmission can be done with less signaling overhead, lower latency and lower power consumption. Information such as UE identity, security information and mobility control information is saved in both UE and network sides when the UE transitions from RRC connected state to inactive state by going through the RRC suspend procedure. The stored information is necessary when UE wants to resume the connection from RRC inactive to RRC connected state for data transmission. With this RRC inactive state, the state transition to RRC connected state becomes more efficient.

Overall, it is beneficial to UE power consumption especially if we consider heavy traffic load or long data packets. For small data packets, the cost, including overhead, latency and power consumption, coming from random access procedure and RRC connection establishment is still too high relative to small data payload [20]. Therefore, introduction of inactive state alone is not sufficient if small data transmission cannot be done under inactive state. This requires enhancements in further steps of 5G evolution.

D. DRX MECHANISM

The data packet arrival is observed to be intermittent. It is power consuming if UE is always required to monitor Physical Downlink Control Channel (PDCCH) for the DL assignment or UL grant. To prolong UE battery lifetime, DRX mechanism was introduced in the early stage of LTE and inherited by NR [9].

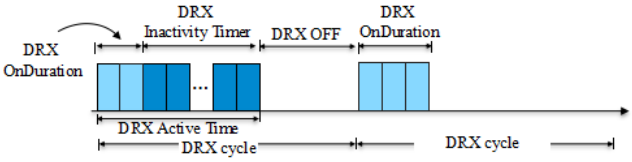

FIGURE 5. DRX mechanism.

If a UE under RRC connected state is configured with DRX, it monitors PDCCH periodically during a configured time duration which is called DRX OnDuration period as shown in Figure 5. If there is no data scheduled, the UE can turn off its RF chain and enter into power saving state. If UE

detects a DCI indicating a new DL or UL transmission, DRX Inactivity Timer would be started. Until the DRX Inactivity Timer expires, the UE needs to keep monitoring PDCCH for the potential subsequent data scheduling.

For a UE under RRC idle/inactive state, the DRX mechanism is related to paging detection. Specifically, the UE needs to detect paging occasion per DRX cycle for paging message and system information update.

The DRX mechanism is a tradeoff between UE power efficiency and data transmission latency. The tradeoff depends on the parameters related to inactivity timer, DRX OnDuration and DRX cycle. Overall, DRX mechanism is beneficial for lowering UE power consumption by allowing UEs to enter power saving mode periodically. There are rooms for further improvement especially considering bursty data. In addition, further enhancements of DRX mechanism for 5G multi-beam millimeter wave communication can be considered, e.g. enhanced beam based DRX measurements [21], directional DRX [22], etc.

E. CONTROL CHANNEL DESIGN

PDCCH Search Space set refers to a set of downlink resources where PDCCH can potentially be carried [23], [24]. To monitor PDCCH, UE performs blind decoding in configured search space set. PDCCH monitoring is a primary contributor to UE power consumption [25]. Another UE power saving feature in Release 15 is that the PDCCH monitoring occasion(s) of a search space set is determined by the parameters such as periodicity(k_s), offset(o_s), duration (T_s) and the monitoring pattern within a slot. Furthermore, the starting symbol for PDCCH monitoring within a slot is represented by a bitmap and the corresponding duration depends on control resource set (CORESET) duration. Specifically, the PDCCH monitoring occasions exist in a slot if the following formula is satisfied and then UE detects PDCCH candidates for consecutive T_s slots.

$$\left(n_f \cdot N_{slot}^{frame,\mu} + n_{s,f}^{\mu} - o_s\right) \bmod k_s = 0 \tag{3}$$

where n_f is the frame number, $N_{slot}^{frame,\mu}$ is the number of slots per frame, $n_{s,f}^{\mu}$ is the slot number within a frame for subcarrier spacing (SCS) μ.

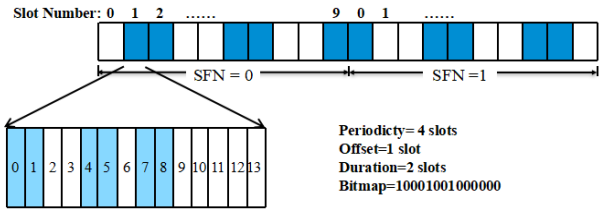

FIGURE 6. PDCCH monitoring occasion(s).

A PDCCH monitoring pattern is illustrated in Figure 6. It is assumed that $k_s = 4$ slot, $o_s = 1$ slot, $T_s = 2$ slots, bitmap = 10001001000000, CORESET duration is 2 OFDM symbols and SCS = 15kHz. Each "1" bit in the bitmap represents a two-symbol CORESET. Hence, symbols 0, 1, 4, 5, 7, 8 are for PDCCH monitoring in the highlighted slots.

The search space set is configured per BWP, UE can adapt to different PDCCH monitoring periodicity through BWP switch. This provides flexibility in some extent but further enhancements to reduce PDCCH monitoring are beneficial especially for the UEs which are not capable of dynamic BWP switching.

F. CROSS-SLOT SCHEDULING

For the slots that are configured with PDCCH monitoring occasions, UE needs to detect PDCCH for data scheduling. It often takes several symbols for UE to finish PDCCH decoding. If Physical Downlink Shared Channel (PDSCH) carrying downlink data scheduled by PDCCH is allocated in the same slot, i.e., same-slot scheduling, UE needs to keep the RF chain on to buffer any potential DL transmission when it is decoding PDCCH.

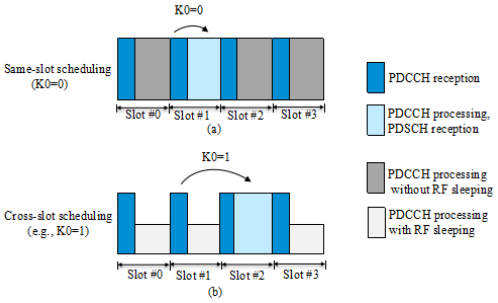

FIGURE 7. Same-slot scheduling and cross-slot scheduling.

As shown in Figure 7, the gap between PDCCH and the scheduled PDSCH is denoted as K0. For the same-slot scheduling in Figure 7(a), UE has to buffer the potential PDSCH for each slot but the DL data is only transmitted in slot#1 Power consumption can be reduced if UE does not buffer PDSCH in slot #0, slot #2 and slot #3. For this purpose, cross-slot scheduling with K0 > 0 is introduced [26], [27]. As shown in Figure 7(b), the PDSCH is assumed to be allocated 1 slot (K0 = 1) after the PDCCH, UE can turn off part of the RF processors and enter into micro sleep after PDCCH is received when there is no PDSCH scheduled in slot #0, slot #1 and slot #3. Compared with same-slot scheduling, the technique of cross-slot scheduling can save a considerable power consumption when the data arrival is sparse. Similar to DL, the time gap (K2) between PDCCH and the corresponding uplink channel, i.e. Physical Uplink Shared Channel (PUSCH) is introduced to reduce UE's UL processing timeline.

III. ENHANCEMENTS FOR POWER SAVING

Based on the standardized framework built in the first release of 5G NR i.e. Release 15, enhancements for power saving have been introduced in Release 16, mainly considering UE

power saving. In this section, UE power consumption model used for the Release 16 study is first provided, and then details of Release 16 enhancements are provided subsequently.

A. POWER CONSUMPTION MODEL

In order to evaluate power saving gain of various techniques, a power consumption model was adopted by 3GPP [25], [28], [29]. The key power states and relative power per slot are provided in Table 5. In addition, three sleep states and the energy consumed by the corresponding state transition are given in Table 6. The relative power is derived from a reference configuration: SCS is 30kHz, one component carrier, 100MHz BWP, 2-symbol PDCCH, the maximum number of PDCCH candidates per slot as 36, PDSCH of maximum data rate (256QAM, 4Rx). Note that the power state of PDCCH-only represents that UE monitors PDCCH without any grant detected.

In the numerical simulation, a single UE is assumed. The simulation focuses on UE behavior on DL initial transmission, i.e., re-transmission and UL transition are not considered. Based on the traffic models and C-DRX configurations provided in Table 7, a time tracing method is used to distinguish the slots with PDCCH monitoring, data reception, or DRX OFF period, etc. The reference configuration and the relative power in Table 5 are used, unless otherwise specified.

To identify the contributors to UE power consumption, the power consumption distribution of different power states is simulated for File Transfer Protocol (FTP) [6] and Voice over Internet Protocol (VoIP) traffic model [30]. The DRX configurations of (DRX cycle OnDuration timer Inactivity timer) = (160ms, 8ms, 100ms) and (4ms, 4ms, 10ms) in Table 7 are used for FTP and VoIP, respectively.

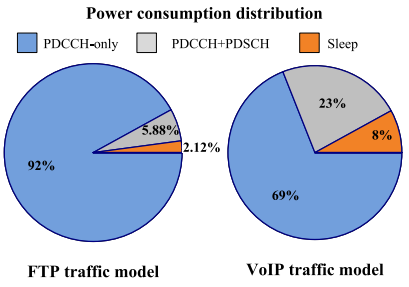

FIGURE 8. Power consumption distribution of different power states for FTP and VoIP traffic model.

According to the simulation results in Figure 8, it can be observed that the PDCCH-only state dominates UE power consumption. Therefore, it is crucial to reduce the power consumed in PDCCH-only state.

B. WAKE UP SIGNALING

As described in Section I, the DRX mechanism allows UE to switch off RF circuit and front-end hardware during DRX off period. However, in the case of sporadic traffic, it is still power consuming for UE to periodically wake up to monitor PDCCH during the DRX OnDuration [31].

Therefore, a wake-up signal indication (WUS) conveyed by DCI signaling with a new DCI format 2_6 [32], which is scrambled by a power saving dedicated identifier, is introduced in Release 16 to inform UE whether or not to start the DRX OnDuration timer for the next DRX cycle for potential data scheduling.

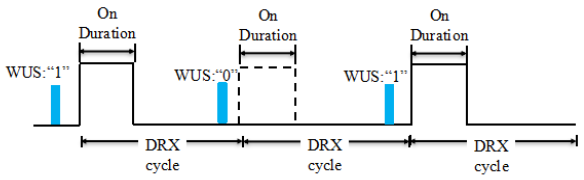

FIGURE 9. DRX mechanism with wake-up indication.

As shown in Figure 9, UE detects the WUS DCI before DRX OnDuration. If the wake-up indication sets to "1", UE should start the DRX OndDuration timer. Otherwise, UE does not need to start the timer.

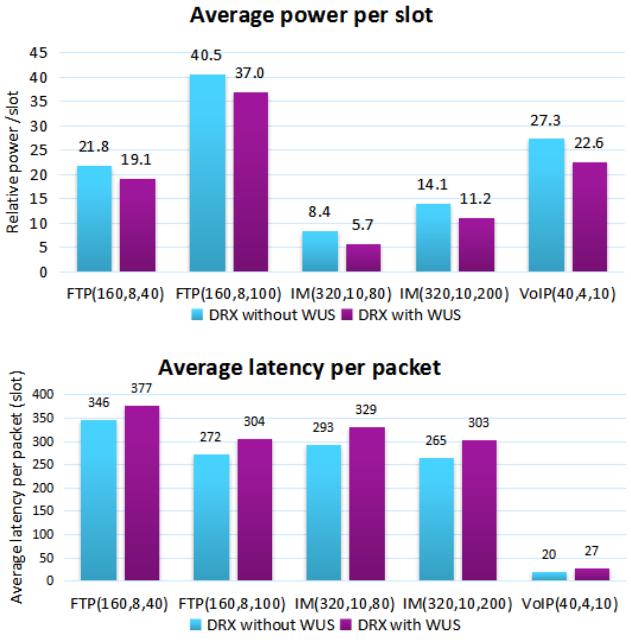

FIGURE 10. Simulation results of DRX mechanism without/with wake-up indication for traffic model (DRX cycle OnDuration timer Inactivity timer).

To evaluate the power saving gain from wake-up indication, the DRX mechanisms without/with wake-up indication are simulated. The results of average power per slot and average latency per packet are shown in Figure 10. The notation of FTP(160, 8, 40) denotes traffic model (DRX cycle in ms OnDuration timer in ms Inactivity timer in ms). The same notation is also applied in the subsequent figures in this paper. It is observed that the mean power of DRX mechanism with wake-up indication can reduce almost 9%-33% power consumption compared with the DRX operation in Release 15 Hence, WUS can provide promising gain on power saving on top of the DRX mechanism. In addition, trade-off between

power saving and latency is studied. It can be observed that latency increase ranges from 9% to 34% with this power saving technique. It is expected that packet throughput would also be impacted due to additional latency.

C. SECONDARY CELL DORMANCY

SCell often refers to frequency carriers in addition to the primary frequency carrier (PCell) by Carrier Aggregation (CA). One of the motivations for SCell is to provide more data bandwidth in other carrier frequencies, typically higher frequencies, for boosting up data throughput while the PCell is more for ensuring the coverage. In Rel-15 Enhancing LTE CA Utilization (euCA), a new SCell state, dormant SCell state, is introduced so that SCells can become inactive for power saving when the data transmission requirement is not high. If SCell is in the dormant state, UE stops monitoring PDCCH in the SCell, but activities such as CSI measurement/reporting and RRM measurement are not impacted. The transition in and out of SCell dormant state are achieved via Medium Access Control (MAC) signaling.

In Rel-16 NR, the dormancy behavior can be implemented at BWP level. The BWP that supports dormancy behavior for the SCell is referred to as dormant BWP, where PDCCH monitoring occasion is not configured. Compared with LTE, the transition between dormancy behavior and non-dormancy behavior is based on BWP switch, which can effectively reduce the switch delay and UE power consumption [33].

The SCell dormancy indication can be conveyed by DCI with different DCI formats for UEs to detect outside and within DRX Active Time. When it is within DRX Active Time, the SCell dormancy indication field is carried by DCI with DCI formats used by data scheduling (i.e. DCI format 0_1 and DCI format 1_1). The information fields that indicate dormancy behavior are in the form of bitmap, where each bit corresponds to one configured SCell group.

When it is outside DRX Active Time, UE detects DCI with DCI format 2_6 on the Primary Cell (PCell) or the Primary Secondary Cell (PSCell) for the SCell dormancy indication. In addition to wake-up indication, the SCell dormancy indication can be carried by the same DL control signaling with WUS using DCI format 2_6. To reduce the blocking rate and resource overhead, DCI format 2_6 can be used to convey information for one or more UEs in a group [34]. Each UE in the group is configured with the location of wake-up indication, while the SCell dormancy indication of that UE is located immediately after the wake-up indication. The structure of the DCI format 2_6 for UE group indication is shown in Figure 11.

One example of SCell dormancy indication is shown in Figure 12. The DCI with DCI format 2_6 detected in PCell outside DRX Active Time indicates that SCell to switch to non-dormant BWP for the data transmission in the subsequent DRX cycle. After data transmission is finished within DRX Active Time, network can send another DCI to indicate UE to switch SCell to dormant BWP to save power.

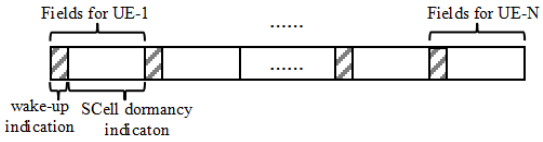

FIGURE 11. Structure of DCI format 2_6.

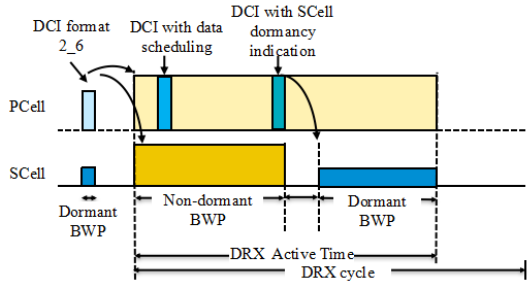

FIGURE 12. Transition between dormant BWP and non-dormant BWP in SCell.

To evaluate the power saving gain from SCell dormancy behavior, it is assumed that two BWPs are configured for both PCell and SCell, where the bandwidths of these two BWPs are 20MHz and 100MHz, respectively In the simulation, UE is indicated to switch to the BWP of 100MHz for PCell and SCell when data packet arrives. After the BWP inactivity timer expires, UE falls back to the BWP of 20MHz. In the cases when SCell dormancy is supported, the BWP of 20MHz is dormant BWP. The results of average power per slot are shown in Figure 13. It is observed that the power saving gain of SCell dormancy behavior under FTP traffic is 18.7%27.4%. As the principles of BWP switching in the two cases are the same, no additional latency is observed when SCell dormancy is supported.

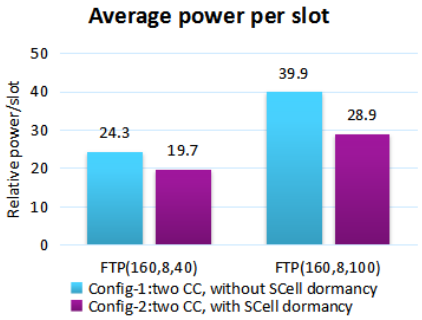

FIGURE 13. Simulation results of two component carriers without/with SCell dormancy behavior.

D. ENHANCEMENT OF CROSS-SLOT SCHEDULING

Based on cross-slot scheduling supported in Release 15, UE can be configured with a list of slot offset K0 values by RRC signaling as described in Section II. A DCI is used to further indicate one K0 value from the configured list for the scheduled PDSCH.

It should be noted that K0 is varied from 0 to 32 slots. If the minimum K0 value configured in the list is larger than zero, it is clear to UE that cross-slot scheduling is applied. However, if the minimum K0 value in the list is equal to 0, PDSCH can be allocated at the same slot with PDCCH. In this situation, UE has to buffer any potential PDSCH before PDCCH is successfully decoded. This would defeat the purpose of power saving by cross-slot scheduling. The situation is the similar for UL.

To guarantee that UE has the knowledge of the minimum slot offset before PDCCH is decoded, a field of minimum scheduling offset restriction is introduced to be carried by data scheduling DCI [34]. When the minimum scheduling offset is applied, UE is not expected to be scheduled with a DCI to receive a data with slot offset smaller than the minimum scheduling offset. The switch between same-slot scheduling and cross-slot scheduling depends on the explicit indication of minimum scheduling offset restriction conveyed by the DCI.

To fully exploit the power saving benefit of cross-slot scheduling, an application delay is introduced before the minimum scheduling offset restriction is updated. The application delay X is determined by

$$X = \max\left(\left\lceil K_{0\min Old} \cdot \frac{2^{\mu_{PDCCH}}}{2^{\mu_{PDSCH}}} \right\rceil, Z_{\mu}\right) \tag{4}$$

where $K_{0minOld}$ is the currently applied minimum scheduling offset of the active DL BWP in the scheduled cell, and Z_{μ} is determined by the subcarrier spacing of the active DL BWP in the scheduling cell and given in Table 8, μ_{PDCCH} and μ_{PDSCH} are the sub-carrier spacing configurations for PDCCH and PDSCH, respectively.

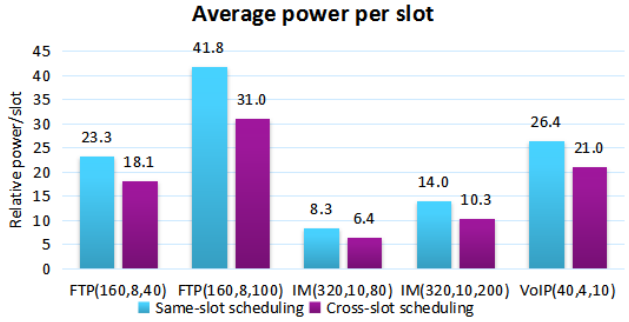

FIGURE 14. Simulation results of same-slot scheduling and cross-slot scheduling.

The results of average power per slot of same-slot scheduling and cross-slot scheduling are shown in Figure 14. For cross-slot scheduling, the additional latency is mainly determined by the value of K_{0min}. To reduce latency, K_{0min} is assumed to be 1 slot in the simulation. The relative power of PDCCH-only for cross-slot scheduling is 70. It is observed that the power saving gain from cross-slot scheduling is 20%27%.

E. MIMO LAYER ADAPTATION

When UE receives signal from base station, it can utilize multiple antennas to achieve receive diversity gain or combining gain for better performance. Meanwhile, the increased number of active antennas consumes more UE energy. As the data traffic arrives in burst, not all the antennas will always be applied. Hence, it would be beneficial to UE energy efficiency if some of the antennas can be indicated to be turned off when small data packet is transmitted or the channel state condition is good enough [32], [35].

In Release 15, the DL maximum number of MIMO layers (Lmax) is configured per serving cell which is common to all the DL BWPs of the serving cell. In Release 16, the DL maximum number of MIMO layers can be separately configured for each DL BWP. The DL maximum number of MIMO layers can be changed through BWP switch which can reduce power consumption by adaptation to less number of receive antennas at UE side.

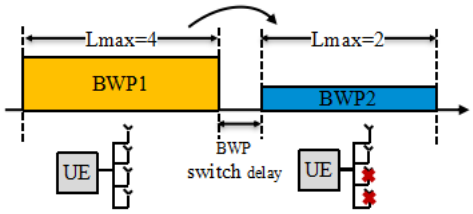

FIGURE 15. DL maximum number of MIMO layers adaptation.

The principle of UE power saving via adaptation to DL maximum number of MIMO layers is shown in Figure 15. In the simulation, one 100MHz-BWP with four MIMO layers is assumed as the baseline. In comparison, two 100MHz-BWPs with four MIMO layers and two MIMO layers respectively are assumed to evaluate the power saving gain from MIMO layer adaptation.

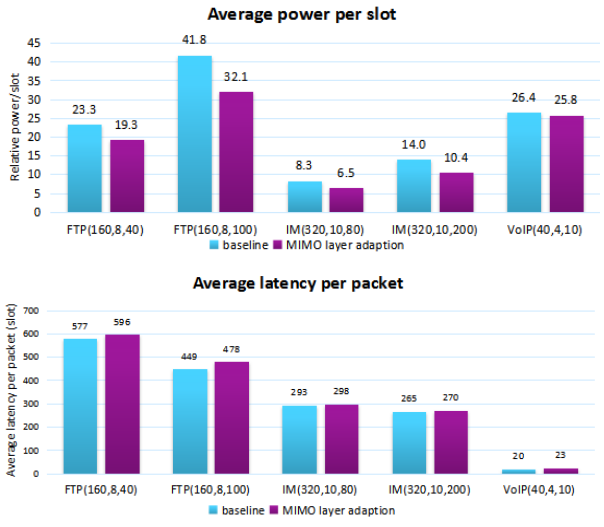

FIGURE 16. Simulation results of without/with MIMO layer adaptation.

The results of average power per slot and average latency per packet are shown in Figure 16. It is observed that the

power saving gain from MIMO layer adaptation is about 2.3%25.2% and the average latency increases by almost 1.7%14.7%.

F. TWO-STEP RANDOM ACCESS PROCEDURE

Two-step random access procedure has been introduced in Release 16 [36]. In addition to reduced signaling overhead and latency, another important benefit of 2-step Random Access Channel (RACH) is power saving. In this section, overview of 2-step RACH and comparison between 2-step RACH and 4-step RACH supported in Release 15 are provided. The procedures of 2-step RACH and 4-step RACH can be illustrated in Figure 17.

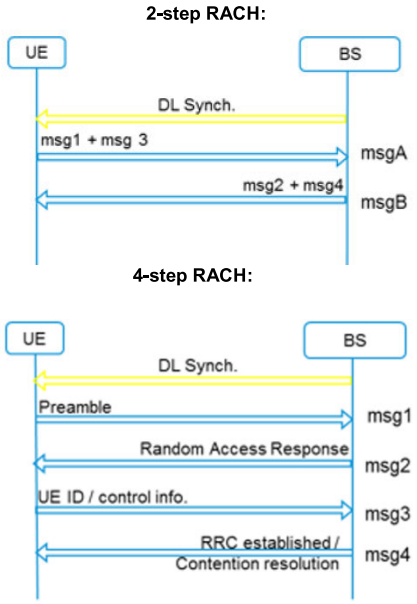

FIGURE 17. RACH procedures for 2-step RACH and 4-step RACH.

2-step RACH can be regarded as a simplified RACH process, where in the first step, Message A (MsgA) contains Msg1 (i.e. preamble carried on PRACH) and Msg3 (can be regarded as payload carried on PUSCH) in 4-step RACH, and in the second step, MsgB contains the Msg2 and Msg4 in 4-step RACH. Hence, for two-step RACH, there is only one time interaction between BS and UE before the establishment of RRC connection for subsequent data transmission.

Frame structure for 2-step RACH and 4-step RACH is referred to Figure 18. Note that a PRB is defined as 12 consecutive subcarriers in frequency domain. The preamble carried on Physical Random Access Channel (PRACH) is used by base station for detecting UE and sometimes for channel estimation if frequency domain interval of preamble can cover that of payload and the time domain offset is not too large. The PUSCH data part can carry some control plane information and potentially carry some small data.

For 4-step RACH, the time and frequency resource for Msg3 is scheduled by base station with dedicated resource. The UL grant for Msg3 is carried on Random Access Response (RAR) sent by base station. While for 2-step RACH, both PRACH and PUSCH resources are shared among a number of UEs for contention-based random access. The set of resources for PUSCH occasion (PO) is periodically configured and associated with the PRACH resource by a predefined mapping rule.

It can be observed that 2-step RACH requires less UE processing compared to 4-step RACH. Hence it has the benefit of power saving especially if a UE is under the scenario with small data traffic which requires the UE to wake up and transmit data intermittently. In such scenario, RACH procedure can consume major part of UE power. Here quantitative analysis is provided where we assume transmission duration of a small data packet is 0.5ms. We consider the following 3 scenarios in Table 2 -4 respectively.

- Scenario1: 4-step RACH + UL data transmission after Msg4
- Scenario2: 2-step RACH + UL data transmission after MsgB
- Scenario3: 2-step RACH + UL data transmission in MsgA

The power consumption calculated in Table 2 -4 is based on the assumption of UE power consumption given in Table A5 in Appendix [20]. It can be observed from Table 2 -4 that about 11% power is saved if uplink data transmission after MsgB in 2-step RACH compared to 4-step RACH. The power reduction mainly comes from having one less reception of Msg2. To achieve more power saving, uplink data transmission can be sent in msgA in inactive state so that it can further skip the reception and processing of dedicated uplink grant and RRC release after finishing the data transmission. As observed from Table 4, about 30% power saving can be achieved if the uplink data is sent in MsgA. In Release 16, power saving can be achieved by introduction of 2-step RACH. However, the gain cannot be fully exploited without allowing UEs to send the data in inactive state i.e. without entering into RRC connected state. Instead of Release 16, the feature of small data transmission in inactive state will be introduced in Release 17.

The analysis above are based on one-shot transmission, i.e. the system behaviors such as decoding failure and retransmission are not considered In realistic system, there could be two cases for MsgA transmission failure, 1) the preamble is mis-detected; 2) preamble is detected but the PUSCH decoding fails. For initial access either with 2-step RACH or 4-step RACH, usually the target mis-detection rate of preamble should be kept as low as possible, e.g. 1%, and thus the additional power consumption caused by mis-detection of preamble is negligible. The second case is more frequently observed, and if happened a fallback mechanism is introduced. When preamble is detected but the PUSCH decoding fails for the MsgA transmission, rather than to retransmit the whole MsgA, the base station can schedule a retransmission of PUSCH only and the rest procedure is the same as 4-step RACH. The retransmission of PUSCH will cause additional

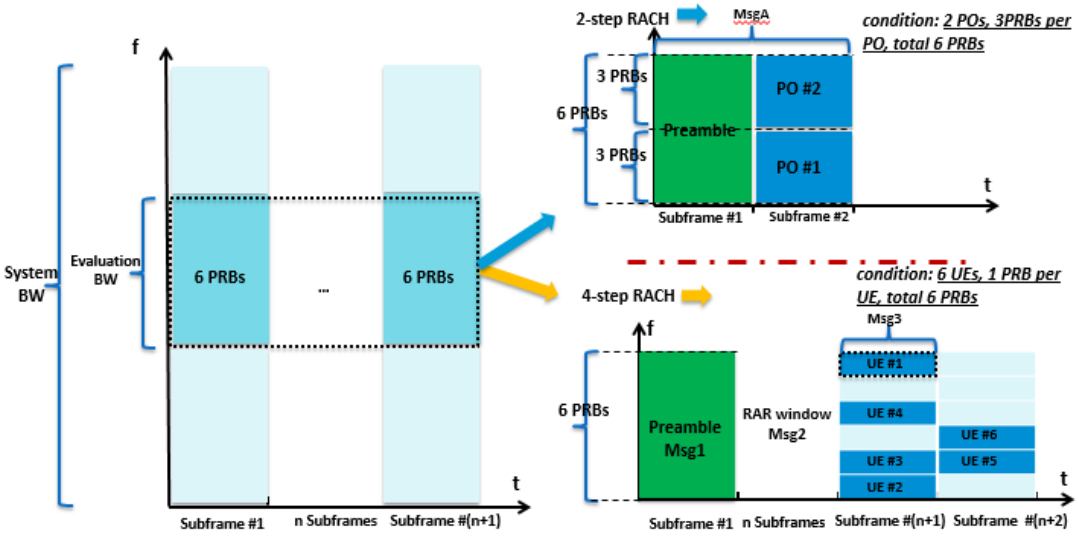

FIGURE 18. Frame structure of two-step RACH and 4-step RACH.

TABLE 2. 4-step RACH + UL data transmission after Msg4.

Message/data	UE Power Consumption (μJ)
Msg1 Preamble transmission (Format A1[9] 0.07ms)	14
Msg2 reception	20
UE processing for Msg2	12.5
Msg3 transmission(0.5ms)	100
Msg4 reception	20
UE processing for Msg4	30
UL-GRANT reception	20
UE processing for UL-GRANT	10
Data transmission(0.5ms)	100
RRC RELEASE reception	20
UE processing for RRC RELEASE	30
SUM	**376.5**

TABLE 3. 2-step RACH + UL data transmission after MsgB.

Message/data	UE Power Consumption (μJ)
MsgA transmission	114
MsgB reception	20
UE processing for MsgB(success)	30
UL-GRANT reception	20
UE processing for UL-GRANT	10
Data transmission	100
RRC RELEASE reception	20
UE processing for RRC RELEASE	30
SUM	**344**

power consumption which is common for 2-step RACH and 4-step RACH. Therefore, it can be imagined that the overall gain of 2-step RACH over 4-step RACH will be diluted if there are multiple retransmissions.

TABLE 4. 2-step RACH + UL data transmission in MsgA.

Message/data	UE Power Consumption (μJ)
MsgA transmission with 0.5ms data	214
MsgB reception	20
UE processing for MsgB(success)	30
SUM	**264**

System level simulation is done to demonstrate the above analysis in terms of the overall gain of power saving versus the traffic load. The packet generation of a UE follows Poisson distribution, and the traffic load is depicted by the packet arrival rate (PAR) at the base station in the unit of packets/s. For the system evaluation, maximum 4 retransmissions (including the initial transmission) are allowed. A cell size of 200m is assumed, the payload size is 72 bits for both MsgA PUSCH and Msg3, and the allocated resource is 6 PRBs and 1 slot. In addition, MMSE-IRC receiver is adopted.

From the evaluation results shown in Figure 19, it can be observed that the overall power saving gain of 2-step RACH over 4-step RACH, i.e. calculated as $(P_{4SR} - P_{2SR})/P_{4SR})$ decreases from 18% to 14.5% as PAR increases from 1 to 6 packets/s. This is because with the PAR increasing, there are more collisions or interferences among UEs and thus more decoding failure can be expected.

G. POWER SAVING TECHNIQUES IN HIGHER LAYER

The power saving techniques involved in higher layer procedure are also considered in Release 16. The power consumption is closely related to UE implementation. It is challenging for network to customize configurations that suit the needs of power saving for all the UEs. To acquire the preferred

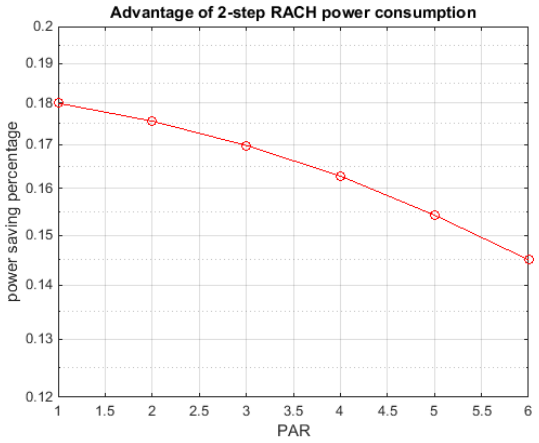

FIGURE 19. Power saving gain of 2-step RACH over 4-step RACH.

configuration at UE side, more UE assistance information is introduced, such as a request of transition from RRC connected mode to RRC idle/inactive mode, minimum scheduling offset values, etc.

In RRC idle/inactive mode, the power consumed by Radio Resource Management (RRM) measurement is significant. To reduce power consumption, neighbor cell RRM measurement can be relaxed based on the evaluation of serving cell for UE with low mobility or at the cell center.

IV. FUTURE TRENDS FOR BEYOND 5G

As the submission of 3GPP 5G solutions for IMT-2020 to ITU has been finalized in June 2019 [37], the standardization process for beyond 5G has been initialized and planned. After two 5G releases, plans for Release 17 enhancements have been mostly finalized [2]. At least the following three topics are related to power saving.

- Support of NR devices with reduced capability, also known as NR Lite UEs [38]
- Small data transmission in RRC inactive state [39]
- UE Power saving enhancements [40]

Regarding NR Lite UEs, it targets at reducing device complexity for UEs applicable to various use cases including industrial wireless sensors, smart wearables and video surveillance. For these new Release 17 UE types, different power saving techniques can be considered compared to typical 5G UEs which have design targets to meet the eMBB and URLLC requirements for in Release 15 and Release 16.

In this section, further trends for beyond 5G power saving enhancements based on new topics are analyzed.

A. SMALL DATA TRANSMISSION IN RRC INACTIVE

Support of Small data transmission in RRC inactive state is one of the planned topics in Release 17 [39]. It targets at traffic with infrequent small data transmission. The exemplary use cases with small data packet include meters or sensors-type NR Lite UEs with periodic measurement reporting, traffic from wearables with periodic positioning information, traffic generated from instant messaging services or heart-beat messages, etc. Under these traffic types, UE is often maintained by network in RRC inactive state. Without support of data transmission in RRC inactive state, UE has to resume the connection with RRC connected state. Connection setup and subsequent release to INACTIVE state happens for each data transmission regardless of how small and infrequent the data packets are. This results in unnecessary power consumption and signaling overhead.

As analyzed in Table 2 -4 in Section III Part F, support of small data in RRC inactive state together with 2-step RACH can achieve significant power saving gain compared to the cases which data has to be transmitted under RRC connected state. In addition to RACH-based scheme, RACH-less scheme will also be supported. Without RACH, small data transmission can be done directly on preconfigured PUSCH resource without preamble. This is expected to provide further power saving gain but this only works under the assumption that transmit timing is known. This applies to the scenarios where UEs are stationary so that synchronization timing acquired in the past can be re-used. For example, industrial sensors are often fixed in the locations with the same indoor environment.

B. UE POWER SAVING IN RRC IDLE/INACTIVE

Given that the major focus in Release 16 is power saving under RRC connected mode, enhancement on the power efficiency of RRC idle/inactive state becomes more urgent considering new use cases for NR Lite UEs. In addition to support of small data transmission in RRC inactive state, power saving technique can be applied to paging and RRM measurement.

The RRC idle/inactive state UE is required to monitor one paging occasion per DRX cycle to detect the scheduling of paging and system information update. The paging occasion location is determined by the UE identification. The false alarm paging rate contributes to the power consumption of RRC idle/inactive state UE, especially in the case of low paging rate.

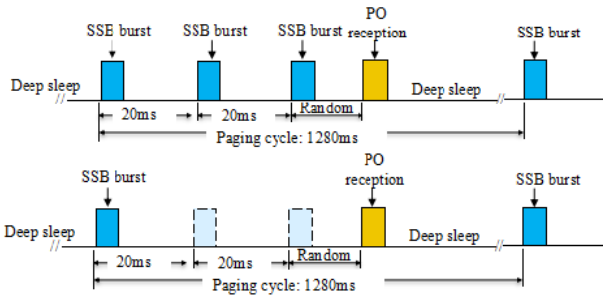

FIGURE 20. UE behavior in RRC idle state.

To ensure the decoding performance of PDCCH and paging message when channel condition is not good enough, UE needs to detect multiple SSBs before PO to adjust automatic control gain (AGC), acquire synchronization in time and frequency domain and RRM measurements, etc. As shown in Figure 20, it is assumed that UE detects three

SSB before the PO. In this case, UE cannot enter into deep sleep, the power consumed by the multiple wake-up times and micro sleep state is significant.

If the number of detected SSB before PO and the paging reception can be reduced, as shown in Figure 20, it enables UE to enter into deep sleep and the power consumption can be significantly decreased.

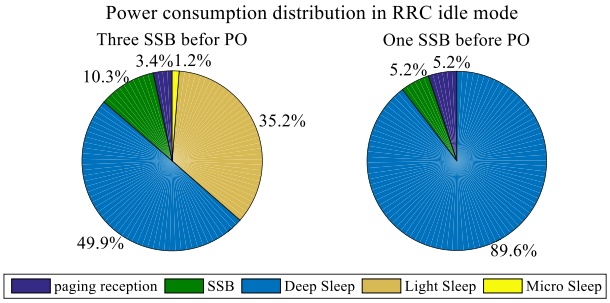

FIGURE 21. UE power consumption distribution in RRC idle state.

In the simulation results provided by Figure 21, the average power per slot is 2.28 and the micro sleep contributes 35% power consumption when the number of detected SSB is 3. When the detected SSB before PO is reduced to 1, the average power per slot is 1.5 and the power saving gain is 34.2%.

To reduce the unnecessary paging reception, solutions such as sub-group paging, group wake up signal (WUS), extended DRX, etc., can be considered.

Meanwhile, UEs under RRC idle/inactive state should perform serving cell measurement and evaluate the cell selection criterion at least once every N DRX cycle, wherein N is determined by DRX cycle. Also, for these UEs, RRM measurement is based on SSB. The periodicity of SSB burst is 20ms by default. If the SSB is not aligned with the paging occasion, UE needs to wake up multiple times to detect the paging occasion and perform RRM measurement. During the time gap between SSB and paging occasion, UE cannot enter into deep sleep, which increases the power consumption. To minimize the gap between the RRM measurement and the paging occasion, additional RS can be considered in Release 17. To reduce the impact on network power efficiency and resource overhead, reference signals, such as CSI-RS, configured to UEs under RRC connected state can be also signaled to UEs under RRC idle/inactive state in broadcast messages. Further, RRM relaxation can be considered for stationary devices.

C. FURTHER UE POWER SAVING IN RRC CONNECTED

In Release 16, wake-up indication can be used to indicate UE to not start the DRX OnDuration Timer when there is no data arrival. However, if UE wakes up for data transmission, the DRX Inactivity Timer would be started and UE has to monitor PDCCH before the timer expiration. Therefore, the following power saving techniques to further reduce PDCCH monitoring within DRX Active Time can be considered in Release 17 [41].

PDCCH Monitoring Adaptation: PDCCH monitoring adaptation allows UE to switch PDCCH monitoring behavior with a sparser PDCCH monitoring occasions within one BWP when data arrives sparsely.

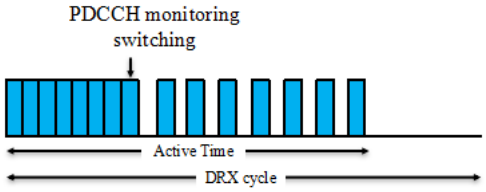

FIGURE 22. UE behavior for exceptional cases of WUS detection.

For example, in Figure 22, UE can be triggered by DCI to switch PDCCH monitoring occasions from ($Ts = 1$ slot, $ks = 1$ slot) to ($Ts = 1$ slot, $ks = 2$ slots), where Ts is PDCCH monitoring duration and ks is the monitoring periodicity.

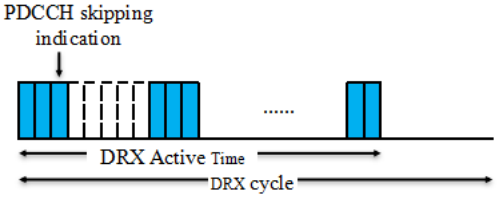

FIGURE 23. PDCCH skipping.

PDCCH Skipping: As shown in Figure 23, with PDCCH skipping technique, base station can send a DCI to indicate UE to perform PDCCH skipping if there is no data to be transmitted to UE. After UE receives the indication, UE can stop monitoring PDCCH to save power.

D. POWER SAVING FOR MASSIVE MIMO

As mentioned in Section I, massive MIMO can provide throughput boost and better coverage in 5G network but the bottleneck still exists in the aspect of power consumption and hardware cost. In this part of section, two future trends on massive MIMO deployment are discussed to address these issues.

Multiple Antenna Panels: Support of Multi-antenna-panel simultaneous transmission is a trend for millimeter wave communications as shown in Figure 23 [42]. In NR beam management and CSI acquisition, multi-panel UE may need to use multiple antenna panels to measure CSI or beam information. The utilization of multiple panels can enable beam sweeping across multiple UE and/or base station panels. In power saving perspective, it is more flexible to turn on/off antenna panel according to the traffic and channel conditions. Multi-panel measurement is not always needed as multi-panel transmission is not always necessary according to the channel variation. Always requiring UE or base station to use multiple antenna panels for beam measurement would cost high power consumption as panel switching. Hence, it is not energy efficient to keep all the panels on for beam measurement.

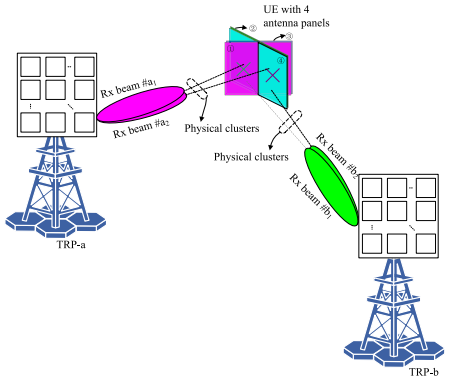

FIGURE 24. Multi-panel transmission.

For the above power consumption aspects, it would be helpful to introduce a mechanism to let UE turn off some antenna ports/panels and keep the panel status aligned with base station. For example, base station can inform UE to activate or de-activate some non-useful ports/panels through dynamic signaling. Further, UE may need to report some information so that base station can make a proper decision considering performance, power consumption and latency. Overall, a standardized panel specific power saving mechanism would be helpful for base station to understand the operating state of UE panels. One example is to apply directional or panel specific DRX mechanism [22].

Intelligent Reflecting Surfaces: Intelligent reflecting surfaces (IRS) is a new promising hardware solution to enhance future wireless communication systems, with consideration of lowering power consumption and enhancing efficiency of massive MIMO networks at the same time. IRS consists of many configurable electromagnetic units (EUs). With the recent developments in meta-surfaces [43], these EUs are low-cost and energy-efficient. These units can be used to radiate or reflect electromagnetic waves, via controlling the electromagnetic properties (e.g. phase, amplitude) in real time [44], [45], adaptive smart beam can be formed at the expected direction(s), achieving coherent superposition at the location of desired receiver, while interference can be kept minimal for the receivers in other locations. With the development of electromagnetic material technology, new types of EUs enabling strong ability and better features have become more practical. Strong ability here means more accurate control of various types of electromagnetic properties, e.g. phase, amplitude, frequency, orbital (OAM) and spin (SAM) angular momentum. Better features include low cost and complexity, thin and light form factors, low power consumption. As long as there are massive number of electromagnetic units and wide distribution, holographic effect can be realized.

IRS proactively modifies the wireless channel via highly controllable and intelligent signal reflection. It can intelligently control the wireless environment, adjust its reflection coefficients dynamically to achieve different functions, e.g. improving signal strength, reducing interference. The signal reflected by IRS can add constructively with those from the other paths to enhance the desired signal power at the receiver. Increase of available paths also make the wireless link more robust and support more spatial layers. IRS based communication is particularly useful when the source and destination have a weak wireless channel in between due to lack of line of sight with obstacles or poor channel conditions. This provides a new degree of freedom to further enhance the wireless link and system performance and paves the way to realizing a smart, programmable and energy-efficient massive MIMO networks. To support this antenna architecture, enhancements on reference signals, CSI feedback and antenna group selection can be considered for the future standard.

TABLE 5. Power consumption models.

Power State	Relative Power
Deep Sleep	1
Light Sleep	20
Micro sleep	45
PDCCH-only(K0=0)	100
PDCCH-only(K0>0)	70
SSB /TRS /CSI-RS	100
PDSCH	280
SSB+PDCCH	170
SSB+PDCCH+PDSCH	300
PDCCH + PDSCH	300

TABLE 6. UE power consumption during state transition.

Sleep type	Additional transition energy: (Relative power x ms)	Total transition time
Deep sleep	450	20 ms
Light sleep	100	6 ms
Micro sleep	0	0 ms

TABLE 7. Configuration of traffic models.

Traffic model	FTP traffic	Instant messaging(IM)	VOIP
Model	FTP model 3 [6]	FTP model 3 [6]	
Packet size	0.5 Mbytes	0.1 Mbytes	As defined in [30].
Mean inter-arrival time	200 ms	2 sec	Assume max two packets bundled.
DRX setting	(DRX cycle, OnDuration timer, Inactivity timer) =(160ms, 8ms, 40ms) or(160ms, 8ms, 100ms)	(DRX cycle, OnDuration timer, Inactivity timer) =(320ms, 10ms, 80ms) or (320ms, 10ms, 200ms)	(DRX cycle, OnDuration timer, Inactivity timer) =(40ms, 4ms, 10ms)

TABLE 8. Definition of Zμ.

μ	Zμ
0	1
1	1
2	2
3	2

TABLE 9. UE power consumption.

Power consumptions	Value (mW)
Power for transmission of uplink channel (PUSCH, PUCCH, preamble)	200(23dBm)
Power for transmit, for non-orthogonal multiple access	100(20dBm)
Power for receiving downlink channel (PDSCH, PDSCH)	40
Connected state (active): - no data transmission/reception - continuously PDCCH monitoring; -Sounding Reference Signal transmission - Measurement - Paging monitoring i.e. it's power-consumption in OnDuration-DRX	48
Connected state (sleep): - no data transmission/reception - no PDCCH monitoring; -no measurement, no Sounding Reference Signal - no paging monitoring i.e. it's power-consumption in inactive-DRX	0.01
Idle state (active): - measurement - paging monitoring	20
Idle state (sleep): Including power consumed for clock and memory maintaining, leakage current.	0.01
UE baseband processing	10

Note: μJ is used to represent the power consumption, where 1μJ =1mW*1ms, 1J=1W*1s.

In addition to the future trends on power saving techniques described from standardization perspective in Part A to Part D of this section, there are some more interesting areas requiring more research and investigation to understand better about the potential impact on standardization and practical implementation. For example, there is on-going discussion [46] on standardization support of Artificial Intelligence (AI) based network power saving which includes information exchange on predicted traffic load between base stations via the standardized interface. Another interesting area is wireless powered communications (WPC) [47], [48] which is especially good for mMTC devices which require low data rate with extensive battery life.

V. CONCLUSION

In this paper, we provide an overview on the standardized framework supporting operations of energy efficient networks in 5G NR. The 5G evolution path on power saving techniques is discussed considering from the first version of 5G standards to future beyond 5G standard development towards 6G. Flexible and scalable system design in 5G NR enables different means of adaptation to various traffic loads and traffic types in the networks. For example, the NR standard supports flexible reference signal design, bandwidth adaptation, DRX mechanism, flexible control channel design and more efficient RRC state transition in Release 15. In Release 16, enhancements on power saving and 2-step RACH have been introduced. Moreover, future trends for beyond 5G evolution on standardization support of enhanced power saving are discussed, taking into account various types of UEs including NR Lite UEs which are potentially introduced in Release 17 with lower complexity targeting at use cases of industrial wireless sensors, smart wearables and video surveillance.

APPENDIX

See Tables 5, 6, 7, 8 and 9.

REFERENCES

[1] *NR and NR-RAN Overall Description; Stage 2 (Release 15), V15.4.0*, document TS 38, 3GPP, Dec. 2018. [Online]. Available: http://ftp.3gpp.org
[2] *Release 17 Package for RAN*, document RAN#86, Dec. 2019. [Online]. Available: https://www.3gpp.org/ftp/Information/presentations/presentations_2019/Rel17_package_RAN.pdf
[3] *Study on Scenarios and Requirements for Next Generation Access Technologies, V14.3.0*, TR 38.913, 3GPP, Jun. 2017. [Online]. Available: http://ftp.3gpp.org
[4] Q. Wu, G. Y. Li, W. Chen, D. W. K. Ng, and R. Schober, "An overview of sustainable green 5G networks," *IEEE Wireless Commun.*, vol. 24, no. 4, pp. 72–80, Aug. 2017.
[5] S. Zhang, Q. Wu, S. Xu, and G. Y. Li, "Fundamental green tradeoffs: Progresses, challenges, and impacts on 5G networks," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 1, pp. 33–56, 1st Quart., 2017.
[6] *Study on New Radio Access Technology—Physical Layer Aspects, V14.1.0*, document TR 38.802, 3GPP, Jun. 2017. [Online]. Available: http://ftp.3gpp.org
[7] Y.-N.-R. Li, J. Li, H. Wu, and W. Zhang, "Energy efficient small cell operation under ultra dense cloud radio access networks," in *Proc. IEEE Globecom Workshops (GC Wkshps)*, Austin, TX, USA, Dec. 2014, pp. 1120–1125.
[8] *NR Physical Channels and Modulation (Release 15), V15.4.0*, document TS 38.211, 3GPP, Dec. 2018. [Online]. Available: http://ftp.3gpp.org
[9] *Medium Access Control (MAC) Protocol Specification (Release 15), V15.8.0*, document TS 38.321, 3GPP, Dec. 2019. [Online]. Available: http://ftp.3gpp.org
[10] Y. Kim, F. Sun, Y. Wang, Y. Qi, J. Lee, Y. Kim, J. Oh, H. Ji, J. Yeo, S. Choi, H. Ryu, H. Noh, and T. Kim, "New radio (NR) and its evolution toward 5G-advanced," *IEEE Wireless Commun.*, vol. 26, no. 3, pp. 2–7, Jun. 2019.
[11] *Adaptation Designs for NR UE Power Saving, RAN1-Ad-Hoc Meeting*, document R1-1900192, RAN1-Ad-Hoc Meeting 1901, 3GPP, MediaTek, Jan. 2019. [Online]. Available: http://ftp.3gpp.org
[12] B. Debaillie, C. Desset, and F. Louagie, "A flexible and future-proof power model for cellular base stations," in *Proc. IEEE Veh. Technol. Conf.*, May 2015, pp. 1–7.
[13] *Base Station Power Model*, document R1-114336, TSG-RAN WG1 #67, NTT DOCOMO, Alcatel-Lucent, Alcatel-Lucent Shanghai Bell, Ericsson, Telecom Italia, San Francisco, CA, USA, Nov. 2011. [Online]. Available: http://ftp.3gpp.org
[14] *Energy Efficiency Analysis of the Reference Systems, Areas of Improvements and Target Breakdown*, document INFSO-ICT-247733, EARTH, Deliverable D2.3, 2010.
[15] *Technical Report-Further Advancements for E-UTRA Physical Layer Aspects, V.9.0.0*, document TR 36.814, 3GPP, Mar. 2010. [Online]. Available: http://ftp.3gpp.org

[16] *NR User Equipment (UE) Radio Transmission and Reception Part 1: Range 1 Standalone (Release 15), V15.4.0*, document TS 38.101-1, 3GPP, Dec. 2018. [Online]. Available: http://ftp.3gpp.org

[17] *NR User Equipment (UE) Radio Transmission and Reception Part 1: Range 2 Standalone (Release 15), V15.4.0*, document TS 38.101-2, 3GPP, Dec. 2018. [Online]. Available: http://ftp.3gpp.org

[18] *Remaining Issue for BWP*, document R1-1806135, RAN1-NR#93, ZTE, 3GPP, May 2018.

[19] I. L. Da Silva, G. Mildh, M. Saily, and S. Hailu, "A novel state model for 5G radio access networks," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC)*, Kuala Lumpur, Malaysia, May 2016, pp. 632–637.

[20] *Quantitative Analysis on UL Data Transmission in Inactive State*, document R2-1701932, RAN2#97, ZTE, 3GPP, Feb. 2017. [Online]. Available: http://ftp.3gpp.org

[21] S. H. Ali Shah, S. Aditya, S. Dutta, C. Slezak, and S. Rangan, "Power efficient discontinuous reception in THz and mmWave wireless systems," in *Proc. IEEE 20th Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Cannes, France, Jul. 2019, pp. 1–5.

[22] M. K. Maheshwari, M. Agiwal, N. Saxena, and A. Roy, "Directional discontinuous reception (DDRX) for mmWave enabled 5G communications," *IEEE Trans. Mobile Comput.*, vol. 18, no. 10, pp. 2330–2343, Oct. 2019.

[23] V. Braun, K. Schober, and E. Tiirola, "5G NR physical downlink control channel: Design, performance and enhancements," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Marrakesh, Morocco, Apr. 2019, pp. 1–6.

[24] F. Hamidi-Sepehr, Y. Kwak, and D. Chatterjee, "5G NR PDCCH: Design and performance," in *Proc. IEEE 5G World Forum (5GWF)*, Silicon Valley, CA, USA, Jul. 2018, pp. 250–255.

[25] *Consideration on UE Power Consumption Model and Preliminary Evaluation Results*, document R1-1812420, RAN1-NR#95, ZTE, 3GPP, Nov. 2018. [Online]. Available: http://ftp.3gpp.org

[26] *NR Physical Layer Procedures for Data (Release 15), V15.4.0*, document TS 38.214, 3GPP, Dec. 2018. [Online]. Available: http://ftp.3gpp.org

[27] *On Adaptation Aspects for NR UE Power Consumption Reduction*, document R1-1812421, RAN1-NR#95, ZTE, 3GPP, Nov. 2018. [Online]. Available: http://ftp.3gpp.org

[28] *Study on User Equipment (UE) Power Saving in NR (Release 16), V16.0.0*, document TR 38.840, 3GPP, Jun. 2019. [Online]. Available: http://ftp.3gpp.org

[29] M. Lauridsen, D. Laselva, F. Frederiksen, and J. Kaikkonen, "5G new radio user equipment power modeling and potential energy savings," in *Proc. IEEE 90th Veh. Technol. Conf. (VTC-Fall)*, Honolulu, HI, USA, Sep. 2019, pp. 1–6.

[30] *LTE Physical Layer Framework for Performance Verification*, document R1-070674, RAN1#48, Orange, China Mobile, KPN, Feb. 2007. [Online]. Available: http://ftp.3gpp.org

[31] *Discussion on Potential Techniques for UE Power Saving*, document R1-1902031, RAN1-NR#96, ZTE, 3GPP, Feb. 2019. [Online]. Available: http://ftp.3gpp.org

[32] *Multiplexing and Channel Coding (Release 16), V16.0.0*, document TS 38.212, 3GPP, Dec. 2019. [Online]. Available: http://ftp.3gpp.org

[33] *Discussion on PDCCH-Based Power Saving Signal*, document R1-1911925, RAN1-NR 99, ZTE, 3GPP, Nov. 2019. [Online]. Available: http://ftp.3gpp.org

[34] *Procedure of Cross-Slot Scheduling Power Saving Techniques*, document R1-1911926, RAN1-NR#99, ZTE, 3GPP, Nov. 2019. [Online]. Available: http://ftp.3gpp.org

[35] *On UE Adaptation to Maximum Number of MIMO Layer*, document R1-1908200, RAN1-NR#98, ZTE, 3GPP, Aug. 2019. [Online]. Available: http://ftp.3gpp.org

[36] *New Work Item: 2-Step RACH for NR*, document RP-182894, RAN#82, ZTE, 3GPP, Dec. 2018. [Online]. Available: http://ftp.3gpp.org

[37] *3GPP Final Technology Submission—Overview of 3GPP 5G Solutions for IMT-2020*, document ITU-R WP5D Contribution 1215, Alliance for Telecommunications Industry Solutions, Jun. 2019.

[38] *New SID on Support of Reduced Capability NR Devices*, document RP-193238, RAN#86, Ericsson, 3GPP, Dec. 2019. [Online]. Available: http://ftp.3gpp.org

[39] *Work Item on NR Small Ldata Transmissions in INACTIVE State*, document RP-193252, RAN#86, ZTE, 3GPP, Dec. 2019. [Online]. Available: http://ftp.3gpp.org

[40] *New WID: UE Power Saving Enhancements*, document RP-193239, RAN#86, MediaTek, 3GPP, Dec. 2019. [Online]. Available: http://ftp.3gpp.org

[41] *Views on Power Saving Enhancement*, document R1-2000513, RAN1-NR#100-e, ZTE, 3GPP, Feb. 2020. [Online]. Available: http://ftp.3gpp.org

[42] Y.-N.-R. Li, B. Gao, X. Zhang, and K. Huang, "Beam management in millimeter-wave communications for 5G and beyond," *IEEE Access*, vol. 8, pp. 13282–13293, 2020.

[43] S. Foo, "Liquid-crystal reconfigurable metasurface reflectors," in *Proc. IEEE Int. Symp. Antennas Propag. USNC/URSI Nat. Radio Sci. Meeting*, Jul. 2017, pp. 2069–2070.

[44] Q. Wu and R. Zhang, "Intelligent reflecting surface enhanced wireless network via joint active and passive beamforming," *IEEE Trans. Wireless Commun.*, vol. 18, no. 11, pp. 5394–5409, Nov. 2019.

[45] S. V. Hum and J. Perruisseau-Carrier, "Reconfigurable reflectarrays and array lenses for dynamic antenna beam control: A review," *IEEE Trans. Antennas Propag.*, vol. 62, no. 1, pp. 183–198, Jan. 2014.

[46] *Summary of Email Discussion on AI-Based Network*, document RP-192579, RAN#86, ZTE, 3GPP, Dec. 2019. [Online]. Available: http://ftp.3gpp.org

[47] K. Huang, C. Zhong, and G. Zhu, "Some new research trends in wirelessly powered communications," *IEEE Wireless Commun.*, vol. 23, no. 2, pp. 19–27, Apr. 2016.

[48] S. Bi, C. K. Ho, and R. Zhang, "Wireless powered communication: Opportunities and challenges," *IEEE Commun. Mag.*, vol. 53, no. 4, pp. 117–125, Apr. 2015.

YU-NGOK RUYUE LI (Member, IEEE) received the B.Eng. degree (Hons.) in electrical and electronics engineering (EEE) from The University of Hong Kong, in 2000, and the M.S. degree in electrical engineering from Stanford University, in 2002.

He is currently a Chief Engineer of the Wireless Product Research and Development Institute, ZTE Corporation. His research focuses on physical layer system algorithms and network optimization in the cellular communication area. He is actively involved in research and standardization activities in wireless communication. He is the Chief Standard Delegate and the Project Manager leading the ZTE Standard Team in 3GPP RAN and RAN WGs. Prior to ZTE, he worked for several telecommunication and semiconductor companies, including Qualcomm and Marvell Semiconductor on the projects related to 2G/3G baseband algorithm design and LTE standardization. His current research interests include power-saving, massive MIMO, mmWave communications, beam management, AI applications on MIMO, URLLC, CoMP, and interference coordination.

MENGZHU CHEN was born in Hubei, China, in 1991. She received the bachelor's degree in optical information science and technology from Wuhan University, Wuhan, China, in 2013, and the master's degree in optical engineering from Tsinghua University, Beijing, China, in 2016. She is currently a Senior Engineer with the Department of Algorithms, ZTE Corporation. She is also a 3GPP RAN1 Delegate. Her current research interests are in the fields of 5G networks, including power saving, channel coding, interference management, URLLC, and AR/VR.

JUN XU was born in Anhui, China, in 1976. He received the bachelor's degree in electromagnetic field and the master's degree in signal and information processing from the Nanjing University of Posts and Telecommunications, Nanjing, China, in 2000 and 2003, respectively.

He has been a Senior Engineer with the Department of Algorithms, ZTE Corporation, since 2003. His current research interests are in the fields of coding, modulation, MIMO, power saving, and link adaptation.

LI TIAN (Member, IEEE) was born in Xiantao, China, in 1988. He received the bachelor's degree in communication engineering and the Ph.D. degree in control science and control engineering from Tongji University, Shanghai, China, in July 2009 and January 2015, respectively.

From 2013 to 2014, he was a Visiting Ph.D. Student at the Department of Electronics and Information Systems (DEIS), University of Bologna. He is currently a Senior Engineer with the Department of Algorithms, ZTE Corporation. His current research interests are in the fields of 5G networks. He is also a 3GPP RAN1 Delegate, as currently the rapporteur of 3GPP Rel.16 Two-Step RACH WI.

Dr. Tian was invited to give a keynote speech at the IEEE CyberC 2019. He also serves as the workshop Co-Chair of the IEEE IWCMC. He serves as a Reviewer for a number of international journals and conferences, including the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, IEEE ACCESS, the IEEE ANTENNAS AND WIRELESS PROPAGATION LETTERS, and the *International Journal of Antennas and Propagation*.

KAIBIN HUANG (Senior Member, IEEE) received the B.Eng. (Hons.) and M.Eng. degrees from the National University of Singapore, and the Ph.D. degree from The University of Texas at Austin (UT Austin), all in electrical engineering.

Since January 2014, he has been an Assistant Professor with the Department of Electrical and Electronics Engineering (EEE), The University of Hong Kong. He used to be a Faculty Member of the Department of Applied Mathematics (AMA), The Hong Kong Polytechnic University (PolyU), and the Department of EEE, Yonsei University. His research interests include MIMO communications, wireless power transfer, and machine learning. He frequently serves on the technical program committees of major IEEE conferences in wireless communications. He was an Elected Member of the SPCOM Technical Committee of the IEEE Signal Processing Society, from 2012 to 2015. He received from IEEE ComSo the 2018 Best Tutorial Paper Award and the 2015 Asia Pacific Outstanding Paper Award, Outstanding Teaching Award from Yonsei, Motorola Partnerships in Research Grant, the University Continuing Fellowship from UT Austin, and Best Paper Awards from IEEE GLOBECOM 2006 and IEEE/CIC ICCC in 2018. He held a University Visiting Scholarship at Kansai University, Japan, in the summer of 2017. Most recently, he served as the Lead Chair of the Wireless Communication Symposium of the IEEE GLOBECOM 2017 and the Communication Theory Symposium of the IEEE GLOBECOM 2014, and the TPC Co-Chair of the IEEE CTW 2013 and IEEE PIMRC 2017. He has edited the JSAC 2015 special issue on communications powered by energy harvesting. He has also served as an Associate Editor of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS (JSAC) series on Green Communications and Networking, and the IEEE WIRELESS COMMUNICATIONS LETTERS. He is currently an Associate Editor of the IEEE TRANSACTIONS ON GREEN COMMUNICATIONS AND NETWORKING.