Pre-Grant Signaling for Energy-Efficient 5G and beyond Mobile Devices: Method and Analysis

Soheil Rostami, Kari Heiska, Oleksandr Puchko, Kari Leppanen and Mikko Valkama, Senior Member, IEEE

Abstract—Due to the severely limited battery capacities, the energy efficiency of mobile devices plays an important role in their usability. In general, the cellular subsystem is one of the major contributors to the energy consumption of a mobile device, thus improving its energy efficiency is of paramount importance. In this article, a new concept of pre-grant message together with associated control plane signaling is introduced, aiming to reduce the energy consumption of the cellular subsystem in the downlink, without notable increase in the buffering delay or latency. The proposed method is fully independent of the ordinary discontinuous reception (DRX) principle, which means that both methods can co-exist and act together to efficiently reduce the energy consumption of the user equipment. The performance of the proposed scheme in terms of the false alarm and misdetection rates are investigated and evaluated, in both AWGN and Rayleigh fading channels. The obtained numerical results show that the pre-grant message signaling can be decoded very reliably and can reduce the system power consumption, relative to an ordinary DRX-only reference system, by up to 70%, 68% and 62% for FTP traffic, video streaming and VoIP, respectively, at the cost of negligible increase in the signaling overhead. The proposed method is also compared in terms of the energy consumption and energy efficiency against another state-of-the-art power-saving mechanism, namely the wake-up radio based approach. The obtained results show that the pre-grant approach outperforms the wake-up based system under broad range of traffic characteristics.

Index Terms—5G, energy efficiency, power saving, discontinuous reception, wake-up scheme, microsleep, signaling, UE.

I. INTRODUCTION

Fifth generation mobile networks, also known as 5G, are expected to provide high data rates and reduced latency in order to deliver a diverse set of new services such as ultra-high-definition video streaming or augmented reality [2]. To satisfy the aggressive requirements of such services, advanced physical layer techniques are vital, however, which commonly result in increased computational complexity and high power consumption on the side of the mobile devices.

Despite mobile devices are continuously becoming faster and more powerful, their batteries still last less than a day under moderate use, making battery life one of the most perennial problems from the devices usability point of view. Specifically, according to a survey by Qualcomm, 60% of the consumers stated that the battery life time and performance are the major attributes that they want to be improved [3]. This issue can be solved either by increasing the energy density of batteries or by reducing the power consumption of the devices. However, mobile devices’ energy density which refers to the battery capacity divided by mobile phone screen area has increased only a few percent during the last decade [4]. Thus, the battery evolution lags far behind the changes and advances in the semiconductor industry.

Modern smartphones have multiple subsystems such as wireless connectivity, an application processor, a graphics processor, audio and video codecs, and human interfaces. Multiple detailed analyses of the energy consumption of commercial smartphones under moderate use show that display, application processor, and cellular subsystem are the main consumers of energy for mobile devices, while the other components contribute substantially only when they are used intensively [3], [5], [6]. Energy consumption of the display and the application processor are out of the scope of this paper, thus we mainly focus on the cellular subsystem and improving its energy efficiency.

In general, the cellular subsystem consists of a transmitter, receiver, and interfaces towards the rest of the terminal. Despite the fact that the transmitter may easily have considerably higher instantaneous power consumption compared to the receiver, it is very important to reduce the energy consumption of mobile devices in the receive mode. The main reason for this is that the data traffic in cellular networks is commonly downlink-dominated, with traffic ratios up to 90% downlink - 10% uplink [7], thus for most of the time, the cellular subsystem operates in the receive mode. Hence, the receiver side energy consumption, when accumulated over a longer period of time, can eventually surpass the transmitter energy consumption.

The 3rd generation partnership project (3GPP) has specified the so-called discontinuous reception (DRX) as one of the key solutions for enhanced battery life [8], by means of switching off the RF circuitry and other modules for long periods. However, despite the adoption of DRX in long-term evolution (LTE) networks, an LTE cellular subsystem is approximately 23 times less power efficient in comparison to WiFi, and even worse compared to 3G systems [9]. In other words, the DRX alone cannot solve the energy consumption challenges in the emerging 5G and beyond systems. The efficiency of DRX is

S. Rostami, K. Heiska, O. Puchko and K. Leppanen are with Huawei Technologies Oy (Finland) Co. Ltd, Helsinki R&D Center, Finland. E-mail: soheil.rostami1@huawei.com, {kariheiska, oleksandr.puchko, kari.leppanen}@gmail.com

S. Rostami and M. Valkama are with Department of Electronics and Communications Engineering, Tampere University of Technology, Finland. E-mail: {soheil.rostami, mikko.e.valkama}@tut.fi

Limited subset of early stage results were presented at IEEE ICC, May 2018 [1].
also known to degrade with different traffic patterns or aperiodic traffic profiles. Furthermore, when optimized for a particular traffic pattern, the DRX increases the control signaling overhead of the network, while also complicating the scheduling [10]. Thus, enhancing the user equipment (UE) energy-efficiency, through alternative or complementary means compared to the basic DRX, is of fundamental importance in 5G systems. We clearly acknowledge that DRX is an existing and standardized energy saving technique, while seek to develop a new and complementary method to further improve the energy efficiency of the mobile devices.

The main contribution of this article is to introduce a novel narrow-band control plane signaling method, referred to as preemptive message (PGM), in order to improve the energy efficiency of mobile devices. Particular emphasis is on traffic scenarios with latency constraints, while DRX is also activated. The idea is to decode the PGM instead of the complete physical downlink control channel (PDCCH) in every subframe of the active period of a DRX cycle, and then switch off energy consuming circuits, if the UE is not scheduled in the current transmit time interval (TTI). With the proposed method, the wasted energy related to empty PDCCH processing and empty subframe buffering are eradicated. The introduced scheme is adapted for the 5G-like frame structure, while can be easily generalized to any other emerging frame structures as well. Furthermore, wake-up radio based energy-saving scheme (WuS) and the proposed PGM are compared in the article, through explicit energy consumption analysis. The obtained results show that the PGM based approach outperforms WuS under broad range of traffic characteristics, without any need for hardware modifications or reconfiguration. We clearly state that the proposed method can, in general, operate and reduce the energy consumption of the mobile device either with or without DRX being enabled. However, since DRX is known to be an existing de-facto energy saving method, we focus in the presentation mostly on the scenario where DRX is enabled.

In this article, vectors are denoted by bold letters, e.g. $\mathbf{H}_x$, while their elements are shown by superscript such that $H_x^{(m)}$ refers to the $m^{th}$ element of $\mathbf{H}_x$. Furthermore, $\hat{\mathbf{H}}_x$ refers to an estimated value of $\mathbf{H}_x$. In addition, $\circ$ and $^\ast$ denote Hadamard product and complex conjugate, respectively. Finally, $\tilde{t}$ is utilized for representing the normalization of absolute time relative to a TTI, i.e. $\tilde{t} = t/TTI$. For readers’ convenience, the most essential variables and notations used throughout the article are collected in Table I.

The rest of the article is organized as follows. Section II briefly reviews the fundamentals of subframe processing, and then describes the reasons behind energy inefficiency of the baseline DRX. Section III shortly reviews and summarizes the state-of-the-art in UE power-saving methods available in literature, while the proposed PGM solution to improve the energy efficiency of DRX is described at concept level in Section IV. The detailed proposed signaling structure of PGM, and the analytical analysis of the involved detection and false alarm rates, are described and provided in Sections V and VI, respectively. An implementation example of PGM receiver using FPGA is briefly addressed in Section VII. Section VIII analyzes and compares the proposed PGM scheme and a reference wake-up radio based scheme, in terms of the relative power saving, followed by system simulation results and their analysis in Section IX. Finally, Section X concludes the work.

### II. BACKGROUND, SYSTEM ASSUMPTIONS AND PROBLEM DESCRIPTION

#### A. Background and System Assumptions

To achieve higher spectral efficiency and peak data rates, orthogonal frequency division multiplexing (OFDM) with a shared channel concept for physical downlink shared channel (PDSCH) is adopted in the 3GPP 5G New Radio (NR) network [11]. Assuming the default configuration with normal cyclic prefix and 15 kHz subcarrier spacing, for presentation simplicity, every 7 OFDM symbols constitute a so-called time slot equal to 0.5 ms, while two consecutive time slots then form one TTI or radio subframe of length 1 ms. In the frequency domain, 12 OFDM subcarriers corresponding to 180 kHz constitutes the basic resource element. A physical resource block (PRB) corresponds then to a time-frequency chunk of 0.5 ms × 12 subcarriers, and is the smallest radio resource unit that is applied, e.g. in scheduling.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>PW_PDSCH</td>
<td>power consumption of PDSCH processing</td>
</tr>
<tr>
<td>PW_PDCCH</td>
<td>power consumption of PDCCH processing</td>
</tr>
<tr>
<td>PW_PGM</td>
<td>power consumption of PGM processing</td>
</tr>
<tr>
<td>PW_WRx</td>
<td>power consumption of the WRx unit in WuS</td>
</tr>
<tr>
<td>PW_sleep</td>
<td>power consumption of cellular module at sleep state</td>
</tr>
<tr>
<td>PW_DRX</td>
<td>power consumption of DRX-enabled cellular module</td>
</tr>
<tr>
<td>PW_WuS</td>
<td>power consumption of WuS-enabled cellular module</td>
</tr>
<tr>
<td>PW_p</td>
<td>power consumption of PGM-enabled cellular module</td>
</tr>
<tr>
<td>TP_PDCCH</td>
<td>processing time of the PDCCH</td>
</tr>
<tr>
<td>TP_PGM</td>
<td>processing time of the PGM</td>
</tr>
<tr>
<td>$\eta_{ip}$</td>
<td>relative power saving of the PGM technique</td>
</tr>
<tr>
<td>$\eta_{sw}$</td>
<td>relative power saving of the WuS technique</td>
</tr>
<tr>
<td>$X$</td>
<td>number of repetitions of PGM symbol</td>
</tr>
<tr>
<td>$M$</td>
<td>code length of orthogonal sequence</td>
</tr>
<tr>
<td>$q_m$</td>
<td>$m^{th}$ orthogonal code samples</td>
</tr>
<tr>
<td>$Y_x$</td>
<td>received OFDM subcarrier samples for PGM repetition index $x$</td>
</tr>
<tr>
<td>$H_x$</td>
<td>channel frequency response at subcarriers corresponding to PGM repetition index $x$</td>
</tr>
<tr>
<td>PW_m</td>
<td>relative power level of the $m^{th}$ orthogonal code</td>
</tr>
<tr>
<td>$p_{md}$</td>
<td>probability of misdetection</td>
</tr>
<tr>
<td>$P_{fa}$</td>
<td>probability of false alarm</td>
</tr>
<tr>
<td>$\beta$</td>
<td>SNR</td>
</tr>
<tr>
<td>$t_m, t_m'$</td>
<td>PGM symbols of $m^{th}$ and $(m + M)^{th}$ mobile devices</td>
</tr>
<tr>
<td>$\nu$</td>
<td>threshold parameter in PGM detection</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>threshold level in PGM detection</td>
</tr>
<tr>
<td>$\Lambda$</td>
<td>test statistic in PGM detection</td>
</tr>
</tbody>
</table>
In order to account for the instantaneous radio channel conditions in terms of fading and interference, and to make efficient use of the channel capacity, link adaptation by means of adaptive modulation and coding (AMC) is employed in LTE and 5G NR. In AMC, the radio-link data rate is controlled by adjusting the modulation and coding scheme (MCS) as a function of the prevailing received signal quality [12]. More specifically, when the radio link conditions are favorable, higher order modulation with higher code rate can be used to increase capacity, while with low received signal quality the modulation order and code rate are decreased. To select an appropriate MCS for the downlink (DL) transmission, UEs report channel state information to 5G base station (called gNB), indicating the instantaneous DL channel quality in both time and frequency domains.

Furthermore, dynamic time-frequency packet scheduling methods are commonly utilized, seeking to provide multi-user diversity gains. Fully dynamic scheduling would be able to exploit multi-user diversity most efficiently but would also require a very large control signaling overhead. That is, each UE needs to be informed about the assigned PRBs in every TTI together with the corresponding AMC parameters. In general, large signaling overheads increase the amount of bandwidth needed for control channels, and hence reduce the bandwidth available for the actual UE data traffic. Therefore, 3GPP has adopted a bandwidth-efficient approach, in which UE is neither informed about its intended downlink control information (DCI) format nor its location within the control region of a subframe. The DCIs are contained in control channel elements, and used to determine the PDSCH’s MCSs and the assigned PRBs. The UE initially receives one of the seven different DCI data blocks on the PDCCH, and then blind decoding is applied, in which UE performs a number of decoding attempts in several control channel elements in each TTI to check if it contains a relevant DCI. The blind decoding complexity and energy consumption depend on the number of decoding attempts of the PDCCH candidate locations for a number of defined DCI formats. Such approach generally reduces the signaling overhead, and therefore increases the capacity of the network, but also simultaneously increases the computational complexity and power consumption of the DCI processing in UE devices [13].

The physical control format indicator channel (PCFICH) and PDCCH are the two main physical channels utilized for DCI. The former indicates the number of OFDM symbols used for the latter in the current TTI. PDCCH as a part of DCI, contains scheduling decisions, required for reception of PDSCH. PCFICH is located in the first OFDM symbol of the subframe, while PDCCH occupies the first one to three OFDM symbols of the subframe. According to LTE and 5G NR frame structures, PDSCH follows immediately after PDCCH reception, and hence, in typical implementations, due to the limits of the maximum processing speed of the baseband unit, UE needs to buffer PDSCH samples, while decoding DCI, ensuring that it does not lose any data. Fig. 1 illustrates how the conventional full-buffering approach operates.

In general, 5G systems are envisioned to provide significantly higher data rates, and lower latency compared to LTE [2]. Therefore, very high channel bandwidths for data transmission are imperative. However, high bandwidth communication can exhaust the mobile devices battery power quickly. It is anticipated that in some use cases, 5G networks may require and adopt bandwidths up to 400 MHz [6], which will impose enormous energy-consumption challenges in the mobile receiver. Moreover, similar to LTE, 5G is an all-IP network, in which UEs frequently have bursty traffic, with occasional periods of transmission activity followed by longer periods of silence. In respect to latency requirements, it is beneficial to scan PDCCH in each subframe to receive DL data traffic or uplink grants, and promptly make an appropriate reaction to the PDCCH [8]. However, this may again largely reduce the battery life time. Therefore, in order to reduce the energy consumption stemming from the continuous monitoring of PDCCH in every subframe, DRX is introduced in LTE and 5G NR, in which UE monitors the PDCCH only during the active periods within a reconfigurable DRX cycle, while switches then off some RF and baseband modules in the corresponding sleep period.

### B. DRX Mechanism and Problem Description

The main parameters required for configuring the sleep and active periods in UE processing are the on-duration timer, inactivity timer, short and long DRX cycles, and short cycle timer [14]. In the active period, the UE is fully functioning and monitors the PDCCH for the duration determined by an on-duration timer, denoted by $t_o$. If a PDCCH message is received during the on-duration period, UE initiates its inactivity timer for the duration of $t_i$. After the inactivity timer is initiated, if a new PDCCH message is received before the expiration of $t_i$, UE re-initiates its inactivity timer. However, if there is no PDCCH message received before expiration of the inactivity timer, a sleep period starts, and UE switches to its sleep state. There are, in general, two different sleep states based on UE power consumption profile, namely, light sleep and deep sleep [14]. When the short DRX cycle, with duration of $t_s$, is employed, UE transitions very often between the sleep and active states, thus the state transition delay needs to be minimum, which in turn can be achieved by selecting the light sleep state. However, when the long DRX cycle, with duration of $t_l$, is utilized, UE may power down more hardware and processing modules to reduce power consumption, through switching to its deep sleep state.
state. However, this commonly results to clearly longer start-up and power-down time windows. Finally, the short cycle timer determines the number of short DRX cycles that UE spends without decoding its PDCCH message, before entering a long DRX cycle. In this work, as a representative number, we assume that the short cycle timer contains 4 DRX cycles.

The aforementioned DRX mechanism and its parameters have, in general, a significant impact on the achievable power savings and the delay incurred by the mobile device applications [15]. According to 3GPP, the DRX mechanisms do not inherently support adapting the DRX configuration and parameters as a function of the UE data applications and the corresponding traffic characteristics. The main reason for this is that such adaptive mechanisms would require largely increased signaling between the UEs and the network. At the same time, however, this makes the DRX latency characteristics suboptimal, and also limits the achievable DRX-based power saving capabilities [16].

In general, depending on the number of active users and the available bandwidth, PDCCH can occupy different numbers of resource elements, ranging from one to three OFDM symbols per subframe. Higher amounts of the PDCCH symbols impact the mobile battery life time adversely through increased energy consumption of blind PDCCH decoding. In addition, as shown in Fig. 1, subframe buffering wastes energy when the UE is not scheduled and thus the PDSCH does not contain data for the UE. As mentioned above, the energy consumption for above type of mechanism increases for higher bandwidths (due to the higher sampling rate and needed higher amount of buffering memory), which thus reduces the energy efficiency of the DRX process.

The average energy consumed by the cellular subsystem in the receive mode during a DRX cycle is dependent on the average time duration that UE decodes DCI, leading to data reception, as well as the average time duration that does not eventually lead to any data reception. According to experimental findings on actual mobile devices available in the market, the time period that mobile device monitors the PDCCH and buffers the subframe without any data allocation has a major impact on battery consumption [4]. Specifically, with DRX, UEs commonly waste 25%, 75%, and 40% of the overall reception time for empty PDCCH processing in order to receive YouTube, Google hangout, and web browsing traffic, respectively [17]. In other words, the main issue with DRX is the large amount of wasted energy in the mobile device during DCI decoding and subframe buffering, while for large share of the time it does not lead to any actual DL data traffic or uplink grants. In 5G systems, due to the large bandwidths, this issue is even more severe than in existing 4G LTE networks.

III. STATE OF THE ART IN ENHANCED POWER SAVING

In this section, two main candidate schemes for enhanced power savings in LTE evolution and 5G mobile devices, namely the microsleep concept and the wake-up scheme (WuS), are reviewed briefly. In general, both methods seek to improve the energy efficiency of the baseline DRX.

A. Microsleep

Authors in [10] and [18] proposed the concept of microsleep, where UE is fully operational for the duration of the first 6 symbols after which it may enter microsleep by switching to the light sleep state for the remaining 8 symbols of the subframe. The concept of microsleep is developed mainly to cope with wasted energy consumption of full buffering of the subframe. Based on this scheme, UE may decode only PCFICH and PDCCH, and then the UE can be powered down within the subframe, if it is not scheduled for data reception in that subframe. However, such scheme requires very fast power switching of various parts of the RF transceiver subsystems, very high-speed processor in the baseband unit as well as large buffering memory. As a result, microsleep increases the cost as well as the instantaneous processing power consumption of the cellular subsystem when fully operational [18]. Furthermore, it is straightforward to show that the microsleep does not impact the buffering delay of the basic DRX, and thus both will incur identical latency.

Numerical results of such microsleep concept in [10] show an aggregate power-saving potential of 5% to 25%, when compared with the full-buffering scheme. However, this comes at the cost of certain degradation in the channel estimation accuracy, since the reference signals of the latter part of the subframe are not utilized [10].

In general, the concept of microsleep can be adopted independently of the DRX. However, its power saving capabilities can be boosted if it is combined with DRX. Fig. 2 illustrates the power consumption profile when utilizing microsleep together with DRX, where it is assumed that the network has data for the target mobile device on the second subframe of the second DRX cycle. As shown in Fig. 2, in case of DRX, the mobile device checks PDCCH every subframe during the on-duration and inactivity timers, and eventually in the second subframe of the second DRX cycle, it realizes that there is upcoming data. For ease of terminology, we refer to such combination of DRX and microsleep in the continuation simply as DRX, and will utilize it as the baseline scheme for comparison purposes.

B. Wake-up Scheme (WuS)

Originally, wake-up based approaches have been utilized in wireless sensor networks (WSNs) to achieve reduced commu-
The microsleep concept does not remove completely the wasted energy of full subframe buffering, since for an empty subframe, the UE still needs to decode and also buffer 6 out of the 14 OFDM symbols with full power consumption. Moreover, this scheme does not remove the wasted energy of empty PDCCH processing at all, which consumes considerable amount of energy. Additionally, microsleep requires hardware upgrades in the RF and baseband processing subsystems to speed up the DCI processing. The WuS, in turn, is very effective approach only when the latency requirements are relaxed, or when the UE receives only very small data payloads. Moreover, the processing time of WRx can be multiple subframes, due to the need for synchronization. In addition, WuS would require new external hardware, which itself increases the costs, size and power consumption of the UE cellular subsystem.

The idea and approach proposed in this article is to reduce the unnecessary energy consumption of the DRX, through the use of PGM and associated signaling which tell the UE whether it is scheduled to receive PDCCH in the next TTI. For the maximum energy savings, the concepts of microsleep-enabled DRX and PGM can also be combined, while terminology-wise we simply utilize the term PGM. In general, the proposed scheme removes wasted energy of PDCCH processing and subframe full-buffering, as well as the need for fast RF subsystem, high-speed processor and large memories for subframe processing required in existing microsleep based DRX. Finally, the processing time of decoding the actual pre-grant message is much lower than that of the full PDCCH decoding or wake-up signal decoding in WuS. In the next section, we introduce the PGM concept more explicitly.

### IV. Proposed Pre-Grant Message Concept

As explained in the previous sections, reducing the energy consumption of empty DRX cycles has, in general, a significant potential to expand the battery life time of mobile devices. This could basically be achieved either by reducing the amount of the empty DRX cycles or the required energy of DCI processing. The former is an inevitable part of the DRX mechanism, and it may or may not be optimized by assigning different DRX configurations per radio bearer [15]. The latter, in turn, could be realized by re-designing the PDCCH in such a way that it requires less energy to decode it. However, such approach would require considerable modifications in the control plane of the existing 3GPP standard.

#### A. Basic Concept

In the proposed method, the mobile device energy consumption is reduced by adopting and transmitting a new narrow-band signal, the PGM, one subframe in advance. The PGM informs of potential scheduling on the PDCCH of the next subframe, or alternatively, if the UE can skip the PDCCH of the next subframe. Because of the signal structure of the proposed PGM, described in detail in the next section, the required energy to decode it is substantially lower than that of the normal PDCCH processing.

In the ordinary full-buffering approach, shown in Fig. 1, the UE processes the full control channels gradually, and is therefore unable to power down. Instead, in the proposed scheme, UE discards decoding the control channels, assuming that the PGM signaling indicates so, and thus saves energy by not buffering...
The proposed approach can also be adopted for radio resource control (RRC) in the idle mode. However, the PGM approach has obviously a larger impact on the mobile device energy consumption in the connected mode, due to the very frequent PDCCH monitoring during connected mode compared to periodical monitoring of the paging channel in the idle mode.

### B. Applicability Aspects

While the largest energy-saving gains can be obtained when the proposed PGM concept is applied together with DRX, it can, in general, be also utilized independent of the DRX by decoding the PGM in every subframe but not entering into deep sleep. In such scenarios, the radio interface latency can be as low as the length of a single subframe, and thus the method is applicable in this form also in ultra-reliable low latency communication (URLLC) applications. Such stand-alone deployment has several advantages, particularly, the flexibility is increased compared to the DRX-based schemes, since there is basically no need for any periodicity in the data traffic patterns. Secondly, the network can send data immediately as soon as there is data available without waiting for the DRX pattern of the UE. Thirdly, it simplifies the radio resource management and UE scheduling in the network. Finally, it fits well with broader range of traffic profiles, without need for any DRX reconfiguration and the associated extra signaling overhead regarding the DRX sleep mode settings. However, the power saving gain of such DRX-disabled PGM scheme is inevitably lower than what can be obtained when the PGM is adopted together with DRX. Thus, in this work, we mainly focus on the combination of the PGM, microsleep and DRX, aiming to reduce the energy consumption as much as possible without notably impacting the buffering delay of the baseline DRX.

The proposed approach can also be adopted for radio resource control (RRC) in the idle mode. However, the PGM approach has obviously a larger impact on the mobile device energy consumption in the connected mode, due to the very frequent PDCCH monitoring during connected mode compared to periodical monitoring of the paging channel in the idle mode.

![Fig. 4. An illustration of UE activity where the PGM of subframe n − 1 has implied the absence of PDCCH in subframe n. Therefore, the UE just decodes the PGM and processes the channel estimation related reference signals during subframe n. After decoding the PGM successfully, the UE goes into microsleep.](image-url)

![Fig. 5. Power consumption profile of the cellular subsystem under the proposed PGM mechanism. In this example, the on-duration and inactivity timers are 3 and 4 subframes, respectively.](image-url)
Therefore, we primarily focus on the connected mode in this article. In general, in the connected mode, each UE operates in either power-active mode or power-saving mode, depicted in Fig. 6. In the power-active mode, the device can receive or transmit in a continuous manner, which is appropriate for URLLC purposes. In the power-saving mode, the mobile device is configured with PGM in order to save battery life time, suitable for eMBB as well as for massive MTC in which large amounts of UEs have small amounts of data at short intervals.

In general, successful completion of the cell search and cell selection procedures as well as acquiring the initial system information and synchronization are required for the UEs in the connected mode. In LTE and 5G NR, the synchronization is acquired by means of the primary and secondary synchronization signals [8]. In this work, we assume that these procedures are correctly executed, and thus the device is operating properly in the connected mode.

V. Proposed Pre-Grant Message Signal Structure

In the proposed PGC concept, each mobile device processes a pre-assigned and pre-reported search space, in time and frequency domains, in order to decide whether it needs to process the actual PDCCH in the next subframe (PGM bit is equal to 1) or not (PGM bit is equal to 0). For this purpose, a time-frequency signal structure, where several PGMs are code multiplexed onto a set of continuous subcarriers, during one OFDM symbol, is applied, such that multiple UEs can be efficiently served in the cell.

In general, the PGM approach requires very low detection error rates in the user equipment. For this purpose, the transmission of the PGM can adopt a power control mechanism reflecting the radio channel conditions of different UEs. In the continuation, the allocated power for the \( m \)th PGM code is represented by \( PW_m \). However, due to the RF power dynamic range restrictions, the desired powers may not always be met, especially for the cell-edge users. Therefore, each PGM is spread on to multiple resource elements to reduce the power differences while at the same time providing the received signal energy necessary for accurate reception and decoding.

The PGM which itself is just a single bit of control information per mobile device is repeated \( X \) times, followed by BPSK modulation and spreading with a length-\( M \) orthogonal sequence, expressed in vector form as \( q_m \) with a dimension of \( M \). Because of the orthogonality among the codes, \( \sum_{m=1}^{M} q^*_m q_m = 0 \), if \( f \neq r \) or equals \( M \), if \( f = r \). In the continuation, we refer to a set of \( 2M \) PGMs transmitted on the same set of \( X \times M \) subcarriers, as a PGM group, and denote by \( x \in \{ 1, ..., X \} \) the repetition index mapping to a specific subset of \( M \) subcarriers. Thus, an individual mobile device can be assigned by a single number (between 1 to \( 2M \)), representing the index of the assigned orthogonal sequence within the group. The first set of \( M \) orthogonal codes are formed by \( M \times M \) Hadamard matrix, and the second set of \( M \) codes are in quadrature \((xj)\) to the first set. After multiplexing the signal representing the PGMs in a group, cell-specific scrambling is applied, which makes neighboring cell interference appear as uncorrelated noise at the UE receivers.

Fig. 7 depicts the signal structure and the generation of the PGM at gNB, together with the basic PGM receiver processing at the UE receiver. For presentation simplicity, other components such as the addition or the removal of the cyclic prefix, FFT/IFFT, and so forth, are not shown.

The deployment of PGM based signaling can cause misdetection and false alarms. In the latter case, the mobile device erroneously identifies 0 as 1 for its decoded PGM bit, leading to unnecessary power consumption in the cellular subsystem. Thus, it is important to control and minimize the probability of false alarms. The former corresponds to the case, where the UE decodes PGM bit as 0 incorrectly, while 1 was actually sent. Such misdetection can add an extra delay, and waste capacity in both PDCCH and PDSCH. Therefore, the misdetection rate requirement of the PGM signaling is eventually stricter than that of the false alarm rate. We will address the detection error rate analysis explicitly in the next section.

In general, the configuration of the PGM parameters is part of the system information, and can be transmitted on the broadcast channel. It contains, e.g., the bits indicating the location of the PGM subcarriers, the scrambling code identifier, and the index number of orthogonal code, together with the DRX-related parameters.

VI. Receiver Processing and Analysis of the Detection Error Rates

After removing the cyclic prefix, demodulating by FFT, and implementing the descrambling process, combined with ideal synchronization assumption, the samples of the received baseband OFDM symbol carrying the PGM information are represented as a vector \( \mathbf{Y}_x \), with dimension of \( M \) for \( x \in \{ 1, ..., X \} \), which can be expressed as

\[
\mathbf{Y}_x = \mathbf{H}_x \otimes (\sum_{m=1}^{M} \frac{PW_m}{2} i_m^*(\mathbf{q}_m) + \sum_{m=1}^{M} \frac{PW_m}{2} i_m^*(\mathbf{q}_m^*) + \mathbf{W}_x), \tag{1}
\]

where \( i_m, i_m^* \in \{-1, +1\} \) for \( m \in \{1, ..., M\} \) are BPSK modulated PGM symbols belonging to the different mobile devices in the PGM group of size \( 2M \). The PGM symbol value of \( i_m = 1 \) informs the UE of potential scheduling on the PDCCH of the next subframe, while the PGM symbol value of \( i_m = -1 \) means that the UE can skip the PDCCH processing of the next subframe. Similarly, \( i_m^* \) is a BPSK-based PGM symbol for the \( (m + M) \)th mobile device in the group, while \( \mathbf{H}_x \) is an \( M \)-dimensional complex vector representing the channel frequency response within the subcarriers corresponding to the subcarrier subset \( x \). Furthermore, \( PW_m \) and \( PW_m^* \) are the relative power
levels of the $m^{th}$ orthogonal code and quadrature of the $m^{th}$ orthogonal code, respectively. Finally, we assume that noise comes from an additive white Gaussian noise (AWGN) process, and can be represented as a circularly symmetric zero-mean complex Gaussian, expressed with an $M$-dimensional complex vector $W_x$, with covariance matrix of $\sigma^2 I$. Without loss of generality, the following analyses in this section are performed from the first ($m = 1$) mobile device perspective.

The likelihood function for decoding the PGM can be expressed as the following test statistic

$$\Lambda = \Re \left( \sum_{x=1}^{X} \sum_{m=1}^{M} q^{(m)}_1(Y_x \circ \hat{H}_x^*(m)) \right),$$

(2)

where $\hat{H}_x$ is the channel estimate for coherent detection, and is acquired based on either the cell-specific reference signals or the demodulation reference signals. The channel estimate can, in general, be written as $\hat{H}_x = H_x + e_x$, where $e_x$ models the estimation error. In this paper, the channel estimation error is assumed to be circularly symmetric zero mean Gaussian with covariance $\sigma^2 I$ [25], [26].

In this work, statistical hypothesis testing is utilized to decide whether to accept $H_0$ or $H_1$. The null hypothesis $H_0$ is a claim that the value of a transmitted PGM equals 0, while the alternate hypothesis $H_1$ is a claim opposite to $H_0$, stating that PGM is 1. The mobile device compares the calculated value of $\Lambda$ from (2) with a fixed detection threshold $\lambda$, and then decides whether PGM is zero or one, as depicted conceptually already in Fig. 7.

By substituting (1) into (2), and ignoring some irrelevant multiplicative and additive terms, $\Lambda$ can be expressed as

$$\Lambda = \sum_{x=1}^{X} \sum_{m=1}^{M} |H_x^{(m)}|^2 \sqrt{\frac{PW}{2}} I_1 + \Re \left( \sum_{x=1}^{X} \sum_{m=1}^{M} H_x^{(m)} W_x^{(m)} \right) + \Re \left( \sum_{x=1}^{X} \sum_{m=1}^{M} e_x^{(m)} W_x^{(m)} \right) + \Re \left( \sum_{x=1}^{X} \sum_{m=1}^{M} H_x^{(m)} e_x^{(m)} \right) + \Im \left( \sum_{x=1}^{X} \sum_{m=1}^{M} H_x^{(m)} e_x^{(m)} \right) \sqrt{\frac{PW}{2}} I_1 - \Im \left( \sum_{x=1}^{X} \sum_{m=1}^{M} H_x^{(m)} e_x^{(m)} \right) \sqrt{\frac{PW}{2}} I_1,$$

(3)

In case of error-free channel estimation, i.e., when $e_x = 0$, $\Lambda$ would have only signal and noise parts (first and second terms). However, due to the presence of unavoidable estimation errors, various undesirable terms show up in the course of calculating $\Lambda$. Such imperfections consist of the self-interference (fourth term), the co-channel interference (fifth term), and some extra noise (third term). When such imperfections are superimposed on the test statistic, some loss in the average SNR is inevitable.

In a simplified case of pure AWGN channel, $\Lambda$ contains the effects of the noise and the equal complex gain of the signal for all subcarriers, i.e., $H_x^{(m)} = h$. The SNR in the AWGN channel case, referred to as $\beta_{AWGN}$, can be expressed as

$$\beta_{AWGN} = \frac{XMPW|h|^4}{\sigma^2 \sigma^2 + \sigma^2 |h|^2 + 0.5 \sigma^2 (PW_1 + PW_1')}.$$

(4)

As can be seen through the above equation, repeating PGM ($X$-times), spreading the data (with length of $M$), and deploying power control ($PW_m$) can increase the SNR with a gain of $X \times M \times PW_m$. Such SNR gain can reduce both the probability of false alarm ($P_{fa}$) and the probability of misdetection ($P_{md}$) over fading channels.

By definition, the probability of false alarm, $P_{fa}$, refers to the probability of deciding to monitor PDCCH, while $H_0$ is true, i.e., $P_{fa} = \Pr(\Lambda > \lambda | H_0)$. The probability of misdetection, $P_{md}$, is in turn the probability of deciding not to monitor PDCCH, while $H_1$ is true, i.e., $P_{md} = \Pr(\Lambda < \lambda | H_1)$. Now, by denoting the inter-distance between the means of $\Lambda$ under $H_0$ and $H_1$ in case of error-free channel estimation by $A = 2\sqrt{|h|^2XMPW_1}$, the aforementioned conditional probabilities under the AWGN assumption can be written as [27]

$$P_{fa,AWGN} = Q(\frac{A + 2\Lambda}{2\sigma_w}),$$

(5)

and

$$P_{md,AWGN} = Q(\frac{A - 2\Lambda}{2\sigma_w}),$$

(6)

where $Q$ is the Q-function representing the tail probability of the standard Gaussian distribution.

In general, the detection threshold $\lambda$ has a direct impact on the achievable detection error rates. An error in a conventional digital communication system is an error, regardless of its more specific type (0 to 1, or 1 to 0). However, the decoding of the
PGM needs to have asymmetric error rates with considerably smaller \( P_{md} \) compared to \( P_{fa} \). Hence, \( \lambda \) should be adjusted accordingly to satisfy the average misdetection and false alarm probability targets.

In this work, for notational convenience yet without loss of generality, we denote the threshold level by \( \lambda = \frac{1}{2} v \), with \( -1 \leq v \leq 1 \). Thus (5) and (6) can be simplified as

\[
P_{fa, AWGN} = Q((1 + v)\sqrt{2\beta_{AWGN}}),
\]

(7)

and

\[
P_{md, AWGN} = Q((1 - v)\sqrt{2\beta_{AWGN}}).
\]

(8)

The actual error probabilities in a slow Rayleigh fading channel can then be evaluated by averaging over the channel realizations as [27]

\[
P = \int_{0}^{\infty} P_{AWGN}(\beta)P_{df}(\beta)d\beta,
\]

(9)

where \( P_{AWGN}(\beta) \) is either \( P_{fa, AWGN} \) or \( P_{md, AWGN} \), both being functions of the SNR \( \beta \). Additionally, \( P_{df}(\beta) \) is the probability density function of \( \beta \), which in case of Rayleigh fading corresponds to a Chi-squared distribution. Thus, through straight-forward manipulations, the results of the integral expressions can be written as

\[
P_{fa} = \frac{1}{2}(1 - \sqrt{1 - \frac{\bar{\beta}}{1 + (1 + v)^2}}),
\]

(10)

and

\[
P_{md} = \frac{1}{2}(1 - \sqrt{1 - \frac{\bar{\beta}}{1 + (1 - v)^2}}),
\]

(11)

where \( \bar{\beta} \) is the average of the SNR over different channel realizations.

Reliability and energy efficiency are both essential characteristics of the PGM-based system, thus a low \( P_{fa} \) and an even lower \( P_{md} \) are preferred. In this work, \( P_{fa} \) and \( P_{md} \) in the order of 10% and 1%, respectively, are assumed and pursued. The target unequal error protections (\( P_{fa} < 10\% \) and \( P_{md} < 1\% \)) can be achieved by configuring the processing gain \((X \times M)\) and the relative powers \( PW_m \) at gNB, as well as the relative detection threshold \( v \) adopted in the terminal detection processing. These can all be controlled by the network operator, as part of the system information, through the broadcast channel.

VII. PGM RECEIVER IMPLEMENTATION PERSPECTIVES

As shown in Fig. 7, the received signal is first descrambled and demapped from \( X \times M \) resource elements in the time-frequency grid. Both processes retrieve back the \( X \times M \) samples, corresponding to the PGM symbols. The resulting samples \( Y_x \) are multiplied with the complex-conjugates of the estimated channel frequency response values, i.e., \( \hat{H}_x \), element by element. Then, the resulting vector \( Y_x \circ \hat{H}_x \) undergoes an inner product with the samples of the vector \( q_k \). After this, as shown in (2), the sum of the inner products \( \sum_{m=1}^{M} q_1^{(m)}(Y_x \circ \hat{H}_x)^{(m)} \) over different values of \( x \) is calculated. At implementation level, for all the involved elements, the needed multiplications are assumed to be performed and the results accumulated as 64 bit values. The real part of the accumulated value is finally taken as the test statistics, as shown in (2), which is thus a 32 bit value. Finally, by comparing to the chosen detection threshold, the mobile device decides on the one-bit PGM value.

This type of PGM processing can be implemented as a sub-module of the overall cellular subsystem. For demonstration purposes, such PGM processor can be designed and implemented, e.g., on a field programmable gate array (FPGA). In this work, the power consumption of implementing the PGM processing on Xilinx Spartan 3E family FPGAs is estimated by using Xilinx power estimator tool. According to the pre-design estimation, the implementation on FPGA can consume up to 60 mW of dynamic power, when the processing gain is \( 4 \times 8 \), i.e., when \( X = 4 \) and \( M = 8 \). Its required clock can be shared from the main baseband unit of the cellular system.

The corresponding device FPGA utilization summary for the proposed PGM processing is provided in Table II. One can observe that even on FPGA, the size and power consumption are very modest and thus reflect good feasibility for actual implementations. Furthermore, while FPGAs are commonly well suitable for reconfigurable demonstrator type of implementations, the involved implementation size and power consumption can be largely reduced, typically by a factor of 7-14, by using application-specific integrated circuits (ASICs) [28].

VIII. PRE-GRANT MESSAGE VS. WAKE-UP SCHEME

Next we pursue explicit analysis and mutual comparison of the power saving capabilities of the proposed PGM scheme and WuS, assuming a Poisson packet arrival model. When it comes to the baseline DRX, the average delay and power consumption where recently analyzed in [14], by using recursive deduction and Markov model. In our analysis, we use the results from [14] as the starting point, and deliberately use similar notations as in [14] for readers’ convenience.

To this end, the probabilities for a UE to be in a short and long DRX cycle are denoted by \( P_{sc} \) and \( P_{lc} \), respectively, and are expressed in closed-form in [14]. Moreover, the probabilities for \( N \) packets to arrive at gNB when the target UE is in a short and long DRX cycle are denoted by \( P_{SN} \) and \( P_{LN} \), respectively, and are expressed in [14]. Additionally, also the probabilities for \( k \) and \( x \) packets to arrive at gNB before and after the end of on-duration period of the target UE in a short and long DRX cycle are denoted by \( Q(k) \) and \( R(k) \), respectively, and are again calculated in [14]. For simplicity, similar to [14], the inactivity timer is not considered in the analysis, and hence UE resides for the durations of \( t_a \) and \( t_s \) at \( t_o \) (or \( t_f - t_o \)) in the active and sleep states, respectively. However, in the numerical results in Section IX, the impact of inactivity timer is also taken into account. Finally, we assume that the number of packets arriving at gNB is maximum one packet per TTI. In scenarios where the number of packets arriving per TTI is more than one, one can
consider multiple packets as one concatenated packet through packet bundling.

A. Relative Power Saving through PGM

We define the relative power saving of the PGM, denoted by $\eta_p$, as

$$\eta_p = \frac{\text{PW}_{\text{DRX}} - \text{PW}_p}{\text{PW}_{\text{DRX}}},$$

where $\text{PW}_{\text{DRX}}$ and $\text{PW}_p$ refer to the average power consumption values of the reference DRX and the proposed PGM scheme, respectively. As shown in Fig. 2 and Fig. 5, the power consumption profiles of the PGM and DRX systems are structurally similar except at the beginning of the empty subframes (i.e., subframes that contain no scheduled data for the UE). Therefore, the numerator can be calculated by considering the difference between the power consumption of PDCCH processing, denoted by $\text{PW}_{\text{PDCCH}}$, and the power consumption of the PGM decoding processing, called $\text{PW}_{\text{PGM}}$, together with the processing times of the PGM ($T_{\text{PGM}}$) and the PDCCH ($T_{\text{PDCCH}}$). Based on this, we can write

$$\text{PW}_{\text{DRX}} - \text{PW}_p = \frac{1}{P_{sc}t_s + P_{lc}t_l} \left( \frac{\tilde{i}_0}{N} \sum_{i=1}^{\tilde{i}_0} \right) + \frac{1}{P_{sc}t_s + P_{lc}t_l} \left( \frac{\tilde{i}_0}{N} \sum_{i=1}^{\tilde{i}_0} \right),$$

$$iQ(i_o - i, N + i - i_o)(P_{\text{PDCCH}}T_{\text{PDCCH}} - P_{\text{PGM}}T_{\text{PGM}}) + P_{lc} \sum_{N=0}^{\tilde{i}_0} \left( \frac{\tilde{i}_0}{N} \sum_{i=1}^{\tilde{i}_0} \right),$$

where the first and the second terms inside the parenthesis correspond to the average power differences, per TTI, between the DRX and PGM caused by short and long DRX cycles, respectively. Additionally, an explicit expression for $\text{PW}_{\text{DRX}}$ is available in [14], in Eq. (32), but not reproduced here for presentation compactness. Hence, $\eta_p$ in (12) can be calculated and evaluated directly by using (13) and Eq. (32) from [14].

B. Relative Power Saving through WuS

Similar to $\eta_p$ in (12), the relative power saving of the WuS, $\eta_w$, is defined as

$$\eta_w = \frac{\text{PW}_{\text{DRX}} - \text{PW}_w}{\text{PW}_{\text{DRX}}},$$

where $\text{PW}_w$ refers to the power consumption of the wake-up based system. In general, an empty DRX cycle occurs when there is no packet arrival during sleep period of the previous DRX cycle or on-duration period of the current DRX cycle. Depending whether operating on the short or long DRX cycles, the probabilities of an empty DRX cycle are different. For the long DRX cycles, the probability of an empty DRX cycle is generally much lower than for the short DRX cycle, under a given traffic model. The overall average probability of an empty DRX cycle, referred to as $P_w$, can be now expressed as

$$P_w = P_{sc} \left( \sum_{k=0}^{\tilde{i}_0} P_{sk}Q(k, 0) \right) \left( \sum_{x=0}^{\tilde{i}_0} P_{sx}Q(0, x) \right) +$$

$$P_{lc} \left( \sum_{k=0}^{\tilde{i}_0} P_{lc}Q(k, 0) \right) \left( \sum_{x=0}^{\tilde{i}_0} P_{lc}Q(0, x) \right),$$

where the first and second terms are the probabilities of receiving zero packets during the sleep period of the previous DRX cycle or the on-duration period of the current DRX cycle, when UE is in short and long DRX cycles, respectively.

The amount of processing power that can be saved during an empty DRX cycle compared to baseline DRX-based system, regardless if it is long or short, is equivalent to the power consumption of a scenario where only on-duration timer is activated, and UE just processes the PDCCH for the on-duration period, namely, $P_w{i_o}T_{\text{PDCCH}}P_{\text{PDCCH}}$. Moreover, in WuS, the WRx adds some extra power consumption to the cellular subsystem. Therefore, the average saved processing power compared to DRX can now be formulated as

$$\text{PW}_{\text{DRX}} - \text{PW}_w = P_w{i_o}T_{\text{PDCCH}}P_{\text{PDCCH}} - \text{PW}_{\text{WRx}} \frac{T_{\text{WRx}}}{P_{sc}t_s + P_{lc}t_l},$$

where $\text{PW}_{\text{WRx}}$ refers to the power consumption of the WRx unit, while $T_{\text{WRx}}$ refers to the WRx processing time, which in this work is assumed to be 14 OFDM symbols [4], [24]. Now, similar to $\eta_w$ in (12), the relative power saving of the WuS in (14), can be calculated by utilizing $\text{PW}_{\text{DRX}}$ in Eq. (32) of [14] and Eq. (16).

C. Numerical Comparison

Next, we show and compare the values of $\eta_w$ and $\eta_p$ for different DRX configuration parameters and packet arrival rates by numerically evaluating the above analytical results. For obtaining the numerical results, the power consumption models used in [14], [24], [29] are utilized, with numerical values as shown in Table III. Notice that the DRX-related power consumption values are needed to evaluate $\text{PW}_{\text{DRX}}$ in Eq. (32) of [14].

The obtained results are collected in Table IV. In general, as can be seen in Table IV, the PGM has higher potential to reduce the energy consumption than WuS ($\eta_p > \eta_w$) for broad range of traffic types and DRX configurations. The reason for the DRX-related power consumption values is higher than that of the WuS.

It can also be observed that for lower packet arrival rate ($\omega = 0.01$ packet per millisecond) under a given DRX configuration, WuS reduces the energy consumption more efficiently than in the case with higher packet arrival rate ($\omega = 0.05$ packet per millisecond). The main reason for such a trend is the presence of a high number of empty DRX cycles for a given DRX setting. However, in case of PGM, the relative
power saving is not strongly dependent on the packet arrival rate, although there is a similar trend observable. This is one of the advantages of the PGM, compared to WuS, implying that PGM is more traffic agnostic and can provide good performance without reconfiguration and re-optimization.

Moreover, increasing the on-duration timer increases the energy consumption of DRX, WuS and PGM due to increasing the duty cycle of the DRX. However, because the probability of empty DRX cycles are not strongly dependent on \(t_o\), \(\eta_p\) does not increase much. In contrast, the probability of empty subframes is much more dependent on \(t_o\), hence \(\eta_p\) improves further with increasing \(t_o\).

Finally, increasing the DRX cycle (short or long), reduces the relative power saving capability of both methods, while simultaneously increasing the buffering delay. The reason for this is that lowering the duty cycle of the DRX leads to less frequent empty DRX cycles and empty subframes. As it can be seen in Table IV, for WuS, reducing long DRX cycle is much more effective than short DRX cycle to improve the relative power saving. This is because at the event of an empty DRX cycle, and when operating under short DRX cycle, WuS saves the wasted energy of PDCCH processing, while in case of long DRX cycle, WuS saves wasted energy of PDCCH processing as well as the start-up and power-down energies of the empty long DRX cycle. However, for PGM, the increase of either long or short DRX cycle has almost similar impact on the relative power saving.

**IX. NUMERICAL RESULTS AND ANALYSIS**

In this section, a set of simulation results are provided and analyzed to validate the proposed PGM concept and its performance, in particular the misdetection and false alarm behavior for different SNR conditions, as well as to show and compare the energy efficiency of the proposed PGM and reference wake-up scheme under practical traffic types.

**TABLE IV**

<table>
<thead>
<tr>
<th>(\omega) [packet/ms]</th>
<th>(t_o) [ms]</th>
<th>(t_s) [ms]</th>
<th>(t_l) [ms]</th>
<th>(\eta_w)</th>
<th>(\eta_p)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.01</td>
<td>2</td>
<td>16</td>
<td>80</td>
<td>32%</td>
<td>50%</td>
</tr>
<tr>
<td>0.01</td>
<td>2</td>
<td>16</td>
<td>160</td>
<td>29.5%</td>
<td>49%</td>
</tr>
<tr>
<td>0.01</td>
<td>2</td>
<td>40</td>
<td>80</td>
<td>31.5%</td>
<td>49.5%</td>
</tr>
<tr>
<td>0.01</td>
<td>2</td>
<td>40</td>
<td>160</td>
<td>29%</td>
<td>48.5%</td>
</tr>
<tr>
<td>0.01</td>
<td>6</td>
<td>16</td>
<td>80</td>
<td>34%</td>
<td>56%</td>
</tr>
<tr>
<td>0.01</td>
<td>6</td>
<td>16</td>
<td>160</td>
<td>32.5%</td>
<td>55%</td>
</tr>
<tr>
<td>0.01</td>
<td>6</td>
<td>40</td>
<td>80</td>
<td>33%</td>
<td>55.5%</td>
</tr>
<tr>
<td>0.01</td>
<td>6</td>
<td>40</td>
<td>160</td>
<td>32%</td>
<td>54%</td>
</tr>
<tr>
<td>0.05</td>
<td>2</td>
<td>16</td>
<td>80</td>
<td>23.5%</td>
<td>45.5%</td>
</tr>
<tr>
<td>0.05</td>
<td>2</td>
<td>16</td>
<td>160</td>
<td>20.5%</td>
<td>44%</td>
</tr>
<tr>
<td>0.05</td>
<td>2</td>
<td>40</td>
<td>80</td>
<td>22.5%</td>
<td>45%</td>
</tr>
<tr>
<td>0.05</td>
<td>2</td>
<td>40</td>
<td>160</td>
<td>19.5%</td>
<td>43.5%</td>
</tr>
<tr>
<td>0.05</td>
<td>6</td>
<td>16</td>
<td>80</td>
<td>27%</td>
<td>54%</td>
</tr>
<tr>
<td>0.05</td>
<td>6</td>
<td>16</td>
<td>160</td>
<td>24%</td>
<td>53%</td>
</tr>
<tr>
<td>0.05</td>
<td>6</td>
<td>40</td>
<td>80</td>
<td>26%</td>
<td>53.5%</td>
</tr>
<tr>
<td>0.05</td>
<td>6</td>
<td>40</td>
<td>160</td>
<td>23.5%</td>
<td>52.5%</td>
</tr>
</tbody>
</table>

**A. PGM Misdetection and False Alarm Simulations**

We consider a single cell which has 16 \((2 \times M)\) DRX-activated UEs using \(M = 8\)-length orthogonal codes, while each PGM is assumed to be repeated either two \((X = 2)\) or four \((X = 4)\) times. Furthermore, each simulation scenario lasts for 10,000 TTIs, and is repeated 100 times over different channel realizations for averaging the results. The multipath intensity profile of the Pedestrian-A channel is assumed. Additionally, an error vector magnitude \((100 \times \sigma_e^2 / P_{ref})\) of 5% is assumed stemming from the reference signal based channel estimation errors. We also assume, for simplicity, that the allocated powers \(P_{W_m}\) for all the UEs are identical, and we vary the received SNR per bit between \(-10\)dB to \(+10\)dB for the PGM channel.

Fig. 8 and Fig. 9 illustrate the PGM detection performance in terms of the false alarm and misdetection rates for two different values of the repetition code length \(X = 8\) for different repetition factors \(X\) and two values of the detection threshold parameter \(\nu\).

![Pfa as a function of SNR per bit with spreading code length of M = 8 for two different repetition factors X and two values of the detection threshold parameter \(\nu\).](image)

**Fig. 8.** \(P_{fa}\) as a function of SNR per bit with spreading code length of \(M = 8\) for two different repetition factors \(X\) and two values of the detection threshold parameter \(\nu\).
Fig. 9. $P_{md}$ as a function of SNR per bit with spreading code length of $M = 8$ for two different repetition factors $X$ and two values of the detection threshold parameter $\nu$.

Fig. 10. $P_f a$ as a function of threshold parameter ($\nu$) with spreading code length of $M = 8$ for two different repetition factors $X$ and two values of the received SNR.

B. Power Saving Evaluations under Realistic Traffic Types

Next, we evaluate the power saving values of the proposed PGM as well as the reference wake-up scheme under more realistic traffic patterns, namely VoIP, video streaming and FTP, compared to the basic Poisson model assumed in the analytical work in Section VIII. In general, the energy consumption of a mobile device in different operating states (sleep, PGM, WRx, PDCCH, and PDSCH processing) is highly dependent on the device implementation and also its operational configurations. For obtaining the numerical results, the power consumption models developed in [14], [24], [29] are again utilized, with numerical values shown already in Table III.

To evaluate numerically the power saving capabilities of the PGM and WuS, we use Matlab simulations in which we implement three different traffic types, namely, video streaming encoded with the H.264/AVC codec (with resolution of 1280x720, codec bitrate of 2433 kbps and frame rate of 30 frames per second), VoIP service with AMR-WB codec known as G.722.2, and FTP traffic. To represent real life video traffic load, we utilize YouTube as our source. The transport protocol is the UDP and the performance of the application depends on the frame size information. In case of VoIP traffic, voice frames are produced using the AMR speech codec, which is a multi-mode codec with 8 speech modes with bit rates between 6.60 and 23.05 kbps. The sampling frequency is 16 kHz with 14-bit resolution and processed at 12.8 kHz. In case of FTP, we consider transferring a 1 GB file. In all cases, we implement the mapping of the source packets into radio subframes, and then calculate the power consumption related to the evaluated scenarios (PGM and WuS), with DRX enabled, by using the assumed power consumption values from Table III. Typical DRX parameters for the corresponding traffic types from [15], [30] are utilized.

The obtained power saving results are collected in Table V at three representative SNR values. As can be observed, similar to the theoretical results in Table IV obtained through a Poisson packet arrival model, both WuS and PGM reduce the power consump-
TABLE V

<table>
<thead>
<tr>
<th>Traffic</th>
<th>$t_0$ [ms]</th>
<th>$t_1$ [ms]</th>
<th>$t_2$ [ms]</th>
<th>$t_3$ [ms]</th>
<th>$E_b/N_0$ [dB]</th>
<th>$\eta_p$</th>
<th>$\eta_w$</th>
</tr>
</thead>
<tbody>
<tr>
<td>VoIP</td>
<td>2</td>
<td>4</td>
<td>20</td>
<td>80</td>
<td>-10</td>
<td>32%</td>
<td>42%</td>
</tr>
<tr>
<td>VoIP</td>
<td>2</td>
<td>4</td>
<td>20</td>
<td>80</td>
<td>0</td>
<td>49%</td>
<td>59%</td>
</tr>
<tr>
<td>Video</td>
<td>4</td>
<td>4</td>
<td>40</td>
<td>128</td>
<td>-10</td>
<td>36%</td>
<td>40%</td>
</tr>
<tr>
<td>Video</td>
<td>4</td>
<td>4</td>
<td>40</td>
<td>128</td>
<td>0</td>
<td>56%</td>
<td>64%</td>
</tr>
<tr>
<td>FTP</td>
<td>10</td>
<td>10</td>
<td>64</td>
<td>160</td>
<td>-10</td>
<td>60%</td>
<td>68%</td>
</tr>
<tr>
<td>FTP</td>
<td>10</td>
<td>10</td>
<td>64</td>
<td>160</td>
<td>0</td>
<td>64%</td>
<td>66%</td>
</tr>
<tr>
<td>FTP</td>
<td>10</td>
<td>10</td>
<td>64</td>
<td>160</td>
<td>10</td>
<td>69%</td>
<td>70%</td>
</tr>
</tbody>
</table>

tion of the basic DRX-based system. However, the PGM clearly outperforms WuS for different considered traffic types and DRX configurations. The performance of PGM and WuS improves further for the FTP packets, due to the existence of more empty DRX cycles and subframes. Furthermore, in these evaluations, the DRX parameters are optimized per traffic, which leads to different on-duration and inactivity timers. In case of FTP, the averages of both timers are higher than with video and VoIP service, leading to larger power saving. Moreover, both methods have similar power saving for FTP traffic type under the given DRX setting. Finally, based on the results, both schemes fit well with FTP traffic pattern, modeling the bursty traffic scenario rather than more periodic traffic patterns.

While the results in Table V focus on three example PGM detection SNR points, the corresponding achievable power saving values for varying SNR values are shown in Fig. 12. The results clearly illustrate that the proposed PGM approach can provide systematically larger power savings than the WuS-based approach, for all the considered traffic types. Additionally, as shown in Table V and Fig. 12, for higher SNR values at which the PGM decoding related $P_{fa}$ and $P_{md}$ values are very low, the achievable power savings are very large regardless of the traffic type. However, even in case of a poor SNR ($-10$ dB), $\eta_p$ in the worst case scenario is 42% which still corresponds to a remarkable energy saving. In general, for a fair comparison, we have assumed that at given SNR, the WRx in wake-up based system has similar false alarm and misdetection rates as the PGM-based system.

X. CONCLUSION AND FUTURE WORK

In this article, methods to reduce the power consumption of DRX-enabled mobile devices during the receive mode without degrading the user experience were pursued, described and analyzed. Novel PGM signaling approach was proposed together with an efficient signal structure and corresponding receiver processing, such that reliable PGM detection, and thus large energy savings can be obtained in 5G and beyond UE receivers. PGM detection performance was quantified analytically, and also its power saving capabilities were analyzed and compared against WuS-based reference approach. It was shown through system simulations that the proposed PGM approach is able to reduce the power consumption of the baseline DRX, by up to 70%, 68% and 62% for FTP traffic, video streaming and VoIP, respectively. Furthermore, the analytical and numerical results verify that the proposed PGM approach outperforms the WuS-based scheme in terms of power consumption for broad range of traffic types without adding any extra buffering delay at the side of the UE devices. In our future work, we will adopt and further develop the PGM concept in standalone mode, without DRX, for URLLC applications. We will also analyze the buffering delay and energy consumption of such a novel stand-alone concept, compared to DRX, under aperiodic traffic patterns including augmented reality and virtual reality.

ACKNOWLEDGMENT

This work has received funding from the European Union’s Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No. 675891 (SCAVENGE), and the Finnish Funding Agency for Technology and Innovation (Tekes) TAKE-5 project.

REFERENCES
