IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received March 26, 2020, accepted April 14, 2020, date of publication April 22, 2020, date of current version May 8, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2989625

Beam-Aware Cross-Layer DRX Design for 5G Millimeter Wave Communication System

AN HUANG[1], (Student Member, IEEE), KUANG-HSUN LIN [1], (Graduate Student Member, IEEE), AND HUNG-YU WEI[2], (Senior Member, IEEE)

[1]Graduate Institute of Communication Engineering , National Taiwan University, Taipei 10617, Taiwan
[2]Department of Electrical Engineering, National Taiwan University, Taipei 10617, Taiwan

Corresponding author: Hung-Yu Wei (hywei@ntu.edu.tw)

This work was supported in part by Qualcomm, and in part by the Ministry of Science and Technology (MOST) of Taiwan under Grant 108-2221-E-002-033-MY3.

ABSTRACT In the fifth-generation (5G) wireless communication system, the Discontinuous Reception (DRX) is indispensable to user equipment (UE) to support evolved technologies. However, the directionality of wireless links, especially in millimeter wave bands, leads the existing DRX mechanism to unnecessary power consumption. In this paper, we first identify the problems of beam pattern mismatching. Then, we proposed a dynamic beam-aware DRX mechanism, which could be applied along with dynamic mmWave beam configuration. The proposed framework jointly optimizes beam-aware DRX operation and dynamic mmWave beam configuration for better energy efficiency. A semi-Markov model is proposed to evaluate a UE sleeping ratio and packet delay. The simulation validates the proposed model. Compared to the baseline 5G NR DRX operation, the proposed scheme achieves 41.6% improvement in the sleep ratio without sacrificing the packet delivery latency performance.

INDEX TERMS DRX, mmWave, dynamic-configured beam pattern, cross-layer design, 5G, beamforming.

I. INTRODUCTION

In the near future, the number of connected user equipment (UE) will increase massively. It brings exponentially increasing global data traffic and exposes considerable power consumption to society. The fifth-generation (5G) New Radio (NR) cellular networks are expected to meet the demands of explosive data, such as 1000x system capacity, 100x energy efficiency, and 10x lower delay [1]. Recently, the network operators are running out of wireless spectrum, so they would like to take advantage of the additional unused resources of millimeter wave (mmWave) frequency bands for data transmission [2]. Therefore, the development in the mmWave technology is essential for the evolution of communication [3], [4]. However, the tremendous propagation loss and sensitivity to blockage in the mmWave band shorten the transmission distance [5]. The deteriorative received power causes low signal-to-interference-plus-noise ratio (SINR) and insufficient throughput.

To increase the received power, a prospective solution is beamforming technology [6]. With beamforming, NR Node B (gNB) could concentrate the transmission power in a particular direction to serve the covered UEs at one time. Connections between gNBs and UEs are established with the beam management procedures [7]. Not only does the beamforming technology enhance the received power of the covered UEs, but it also reduces the impact of interference between the different gNBs [8]. However, the massive antenna array used in mmWave beamforming technology consumes higher power than the antenna of the traditional devices and therefore reduces the battery life on UE side [9].

Discontinuous Reception (DRX) is a crucial technique for power saving. With DRX, a UE can turn off its radio-frequency (RF) module properly, and decrease power consumption without yielding unacceptable delay. It is expected that UEs in 5G networks will continue using DRX mechanism to save power [10]. Nevertheless, employing DRX in mmWave bands without considering the directionality of mmWave link will lead to sub-optimal performance. Therefore, the beam-aware Media Access Control (MAC) Layer protocol design is essential to reduce the energy consumption of UEs in the mmWave 5G wireless networks [11]–[13]. The gNB could only serve part of the UEs at the same time. Thus, the gNB needs to serve the UEs on different beams

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

sequentially, and the order of the beams could be regarded as the serving beam pattern. If the UEs could operate their RF modules with the consideration of serving beam patterns, the amount of additional reduced power consumption is considerable, especially in the 5G NR networks with a massive number of mobile devices.

In this paper, we propose a cross-layer design that enables the DRX mechanism adaptive to the beam patterns in mmWave communications. Since the order of serving beams influences the performance of the DRX mechanism, the performance could be further improved with the assistance of the suitable serving beam patterns. The main contributions of this work are summarized as follows:

- With the awareness of beam patterns, we design a DRX mechanism that is adaptive to the mmWave beamforming technology. Most research on 5G NR DRX focuses on integrating the beam searching process into the DRX mechanism, or improving the DRX mechanism under different scenarios in 5G networks [14]–[20]. To our best knowledge, our DRX mechanism is the *first* joint design of the DRX mechanism and the discrete serving beam pattern. (§V)
- We propose the models of the static and dynamic serving beam patterns in the future 5G networks. Considering the various requirements and scenarios, we describe how the gNB configures DRX mechanism with these beam patterns, and we show the tradeoffs between these two options. (§IV)
- We analyze the performance of the proposed cross-layer DRX mechanism via a semi-Markov model. The expected values of the main performance metrics, *i.e.*, sleep ratio and packet delay, match the simulation results. (§VI and §VII)
- We explore the impact of the serving beam pattern on the proposed cross-layer DRX. The gNB can select beams to schedule data flexibly in a specific period. The flexible scheduling period can enhance both the sleep ratio and the packet delay of the proposed DRX mechanism. (§VIII)

II. BACKGROUND AND RELATED WORK

A. DRX IN 5G NR NETWORKS

The DRX mechanism in 5G NR follows that in Long Term Evolution (LTE) and can support different numerologies with the timers specified in milliseconds [10], [21]. As the example in Fig. 1, a DRX cycle contains one *ON Duration* and one *Opportunity for DRX*, in which a UE turns its RF module on and off, respectively. In *ON Duration*, the UE is required to monitor Physical Downlink Control Channel (PDCCH) continuously. Any indicators for packets make the UE reset their *drx-Inactivity Timer* and extend the wake-up period. If *drx-Inactivity Timer* expires, the UE will turn off its RF module and enters *Opportunity for DRX*, in which it is allowed to turn off its radio to reduce the power consumption but at the expense of ceasing detecting the indication of data at

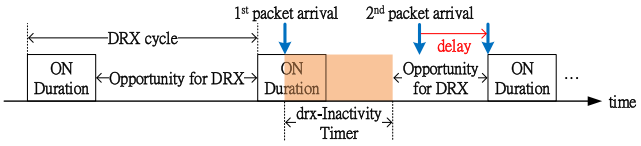


FIGURE 1. The DRX operation in the 5G NR cellular network. A DRX cycle consists of an *ON Duration* and an *Opportunity for DRX*. Each packet reception on UE side would make the UE reset its *drx-Inactivity Timer* and extend the *ON Duration* time. The packet arrived in the *Opportunity for DRX* are buffered by gNB until the next *ON Duration*.

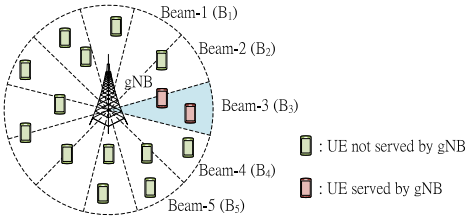


FIGURE 2. An illustration of directional wireless communication with 12 beams on gNB side. The gNB serves the UEs on one beam in each beam period. In this example, two UEs may be served in the beam period when the gNB uses Beam-3 to perform downlink transmission. The UEs on different beams can not receive the downlink signal.

that time. The packets arriving in *Opportunity for DRX* are buffered until the UE enters *ON Duration*, so the packet delay increases.

B. BEAMFORMING TECHNOLOGY AND THE CHALLENGES

For the lack of bandwidth resources, mmWave communication is considered as a potential technology. On the one hand, the mmWave communication relieves the spectrum shortage as well as satisfies the requirement for a higher data rate. On the other hand, the tremendous propagation loss of mmWave weakens the received power. The gNB concentrates the transmission power in specific directions to enhance the received SINR in the 5G NR communication networks. Via beamforming technology, the received power level increases. This technology could benefit not only the communication in lower frequency bands but also the one in mmWave bands.

Owing to the spatial locations of the UEs, they are served with different beams (as shown in Fig. 2). The gNB configures the DRX parameters of every UE and knows when the UEs are active. When there are buffered packets and the destination UE is active, the gNB can schedule data on the correct beam. The performance of the NR DRX under 5G NR networks is almost the same as that in the LTE networks, despite there is room for improvement in energy efficiency. Fig. 3 illustrates the unnecessary UE power consumption in 5G networks. The gNB could only select limited beams to schedule data at one time. In this case, we assume that the gNB uses one beam for transmission at one time. In every time-slot, the gNB determines the beam to be used for downlink transmission, but the UEs do not have the information of the selected beam. It is power-consuming for the UE on Beam-3 to turn on the radio and extend its *ON Duration* while the gNB is serving other UEs on another beam. The

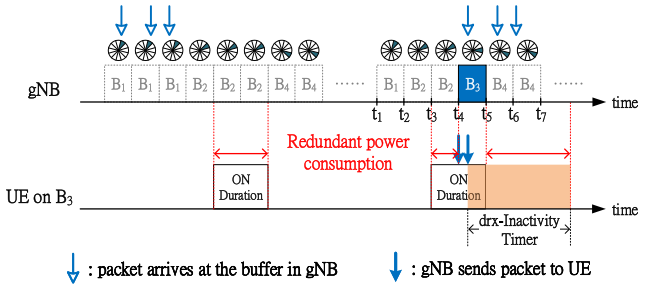


FIGURE 3. An illustrative example of the unnecessary power consumption problem in 5G NR DRX operation. In the first *ON Duration* of the UE on Beam-3, the UE turns on its RF module while the gNB is using another beam to perform downlink transmission. The second *ON Duration* is prolonged after the packet reception, but the gNB is not serving the UEs on Beam-3 after time-slot t_5. In both cases, although there are buffered packets for the UE, the gNB is not able to perform the downlink transmission for the UE.

energy inefficiency should not be omitted, particularly for the scenarios with a massive number of UEs. Therefore, we aim to propose a new DRX design to decrease the power consumption without a larger packet delay for 5G NR networks.

C. RELATED WORK

Researchers have investigated power saving mechanisms in 5G mmWave communications. Lauridsen *et al.* combined microsleep, DRX, Discontinuous Transmission (DTX), and an aperiodic sleep mode with a wake-up signal receiver. In the short and pipelined 5G frame structure, they showed that the battery life of the 5G mobile devices could be improved from 20% to 90% compared to LTE [22]. There is also research focusing on the impact of the beamforming technology on the DRX mechanism. In [11], Beam-Aware DRX was proposed with the consideration of periodic serving beam cycles, and the performance was evaluated via semi-Markov Chain analysis. Considering the device-to-device (D2D) unmanned aerial vehicles (UAV) mmWave communication framework, Zhang *et al.* proposed a fast beam tracking DRX mechanism to overcome the easily-occurred beam misalignment due to the high mobility of UAVs [20]. The state-of-the-art approach is Directional DRX (DDRX) proposed by Maheshwari *et al.* [16]. To maintain the best beam pairs, a four-state DRX mechanism was proposed to be integrated with the process of beam searching. In [15], [16], the UE with Standalone DDRX (S-DDRX) operates beam searching after every DRX cycle.

For the DDRX mechanism, there are extensions under different scenarios to improve Quality of Service (QoS) or power efficiency. With the consideration of dual connectivity (DC) of the UE, the authors proposed Integrated DDRX (I-DDRX) in [14], [16]. The UEs can detect the indication of data via LTE link, so they only need to search for the best beam pair when packets arrive. To avoid the overhead for maintaining DC and unnecessary beam searching procedure, Liu *et al.* proposed a DRX mechanism for multiple beams, with which the UEs only operate beam training when they monitor poor channel quality instead of the indicator of downlink data [19]. In the Standalone D2D communication networks, two UEs with Cooperative DDRX take turns searching for the best beam pair and inform each other [16]. In the 5G networks, various service requirements are attainable with the flexible transmission time interval (TTI). With the consideration of delay constraints, Maheshwari *et al.* proposed an algorithm to find a suitable TTI length in S-DDRX [17]. To reduce more power consumption in the light traffic with the short bursty data, Philip *et al.* improved the design of I-DDRX by adding an auxiliary active state after each DRX cycle when packet arrival number is small [18].

Compared to the previous research, we focus on cross-layer design and optimization between MAC layer DRX and Physical layer mmWave beam pattern configuration. In MAC layer, DRX is aware of mmWave beam patterns so that the system could operate in an energy-efficient way. In Physical layer, a dynamic beam pattern configuration framework is proposed, and cross-layer optimization is conducted to handle the heterogeneous device and traffic distribution. To the authors' best knowledge, this research work is the first cross-layer framework that jointly optimizes beam-aware DRX operation and dynamic mmWave beam configuration.

III. PROBLEM DESCRIPTION

As shown in Fig. 3, the 5G NR DRX leads to inefficiency in power consumption due to the lack of information about beam directionality. The UE on Beam-3 turns on its RF module in *ON Duration*. The packet arrival makes the UE reset its *drx-Inactivity Timer*, and the wake-up time of the UE is extended. The gNB uses Beam-3 to serve the UE only in some specific periods. That is to say, there is no need for the UE to monitor PDCCH continuously. In this work, we proposed a 5G NR DRX framework to solve such a problem of beam behavior asynchrony. The two main proposed designs on gNB side and UE side are shown as follows:

- We will define the two main strategies of the beam serving pattern on gNB side. In the proposed serving beam pattern, the difference between two strategies and what the gNB and the UE behave are introduced in §IV.
- We will demonstrate the proposed 5G NR DRX design, which is adaptive to the serving beam pattern in §V, and show how to use a semi-Markov Chain to model the DRX operation in §VI.

IV. BEAM PATTERN IN 5G NR NETWORKS

A. THE DEFINITION OF BEAM PATTERN

In the mmWave communication networks with beamforming technology, the beam pattern for scheduling data plays an important role. Intuitively, there are two options for the beam patterns. One is repeating the same beam pattern to serve periodically, which is called *"Static-Configured Beam Frame"* in this paper; the other is selecting beams in the arbitrary order, and we describe this option as *"Dynamic-Configured Beam Frame"*. The brief descriptions of these two beam patterns are as follows:

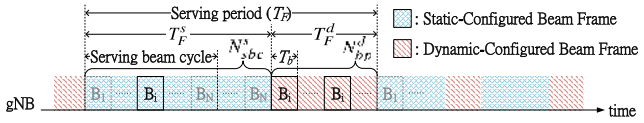


FIGURE 4. The proposed structure of the serving beam pattern is based on the serving period which consists of a Static-Configured Beam Frame and a Dynamic-Configured Beam Frame. In Static-Configured Beam Frame, the gNB serves the beams sequentially. In Dynamic-Configured Beam Frame, the gNB dynamically decides the serving beams.

1) **Static-Configured Beam Frame:** Fig. 4 illustrates that the gNB schedules the UEs on the same beam in each time-slot, and sequentially serves all beams in one serving beam cycle. If there are downlink traffics for the UEs on the beams other than the serving beam, these packets are temporarily buffered. Ideally, the maximal delay would not exceed one cycle.
2) **Dynamic-Configured Beam Frame:** In the Dynamic-Configured Beam Frame, the gNB could select beams to serve the UEs in urgent need dynamically. Some specific applications require extremely low delay, such as video streaming services and Ultra-Reliable Low-delay Communication. Besides, the uneven spatial distribution of UEs in various scenarios causes diverse data traffic demands. Under such scenarios, the fixed resource allocation in the Static-Configured Beam Frame is improper, so beam frames with the flexibility of scheduling are needed. Since the gNB dynamically decides the serving beams in Dynamic-Configured Beam Frame, the UEs are unable to predict the exact serving beam pattern. Therefore, mechanisms to coordinate the gNB and UEs are essential, or all UEs need to stay awake in Dynamic-Configured Beam Frame.

B. THE PROPOSED SERVING BEAM PATTERN

We would like to design a DRX mechanism in the 5G NR networks to improve the power efficiency for the UEs with prior knowledge of the best beam pairs. As mentioned above, the Static-Configured Beam Frame and Dynamic-Configured Beam Frame are the two main strategies to determine the serving beam pattern for gNB. It is essential to consider the massive effect of serving beam pattern on the DRX operation when we evaluate the performance. In this paper, we propose the 5G NR beam pattern demonstrated in Fig. 4. There are one Static-Configured Beam Frame (T_F^s) and one Dynamic-Configured Beam Frame (T_F^d) in a serving period ($T_F = T_F^s + T_F^d$), and the Static-Configured Beam Frame consists of N_{sbc}^s serving beam cycles. Besides, only when the gNB may serve the UEs do they turn on radios to monitor PDCCH. Thus, the UEs could decrease unnecessary power consumption in the Static-Configured Beam Frame.

In the Dynamic-Configured Beam Frame, the gNB tends to serve the UE in urgent need of data transmission. We assume that the UEs with more history traffic have higher probabilities of requesting for the wireless resources. Therefore, the gNB decides whether to serve the UE or not based on the number of arriving packets during the interval (t_{begin}^s, t_{ls}^s), where t_{begin}^s is the beginning of Static-Configured Beam Frame and t_{ls}^s is the time when the gNB last serves the UE in Static-Configured Beam Frame. If at least one UE is receiving no less than $N_{pkt,th}$ packets on the beam (*i.e.*, the beam *achieving the "serving requirement"*), the gNB will add the beam to the serving candidate list. According to the total number of packets arriving on different beams, the gNB selects which beams to schedule data from the serving candidate list in the following Dynamic-Configured Beam Frame randomly. When multiple beams *achieve the "serving requirement"* simultaneously, the gNB might not serve all the candidate beams because of the limited length of Dynamic-Configured Beam Frame. From the UE's perspective, if the received packet number of the UE is more than or equal to the threshold ($N_{pkt,th}$), it will turn on its RF module to detect the indicator of data in Dynamic-Configured Beam Frame. In this paper, we refer to the event as **"Matching"** that the gNB serves the UE in the Dynamic-Configured Beam Frame when the UE turns on its RF module in the Dynamic-Configured Beam Frame.

V. THE PROPOSED DRX MECHANISM IN mmWave COMMUNICATION

The design goal in the 5G NR DRX operation is to deliver data traffic in an energy-efficient manner. In consideration of beamforming technology, we propose a cross-layer DRX design that is adaptive to the Static-Configured and Dynamic-Configured Beam Frame in the mmWave communication networks. Fig. 5 illustrates the proposed cross-layer DRX operation. It consists of four states, including *Active*, *Inactive*, DRX_{check}, and DRX_{sleep}, with the consideration of the beam pattern. *Active* and *Inactive* correspond to the *ON Duration* and the *drx-Inactivity Timer* in the NR DRX respectively. When there are packet arrivals, the UE stays in *Active* or *Inactive* to detect the indicator of downlink data and respond immediately. The UE with the proposed DRX mechanism turns on its RF module according to the serving beam patterns. The DRX cycle includes one DRX_{check} and one DRX_{sleep}. To enhance the energy efficiency, the UE wakes up for one beam period (T_b) to detect the buffered traffic in DRX_{check}, and then go to sleep in DRX_{sleep}. In DRX_{check}, any packet arrival makes the UE enter *Active*. In the following paragraph, we will describe how to handle the energy inefficiency problem mentioned in §III.

The beam-aware DRX configuration: In the proposed DRX mechanism, the gNB could configure the DRX operation of the UEs in terms of serving beam patterns. With the detailed information of the beam pattern, the UEs turn on their radio only when the gNB is serving them. To ensure that the UEs could control their RF modules with the awareness of serving beam patterns, the proposed DRX mechanism includes the following two settings. First, for each UE, the gNB configures the duration of *Active*, *Inactive*, and DRX cycle as multiple of serving period (T_F). Second, the gNB informs the UEs of the beam pattern configuration. With

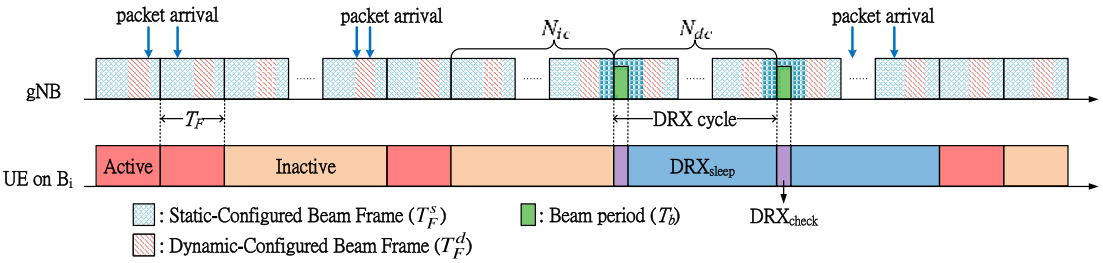


FIGURE 5. The proposed cross-layer DRX operation in the Static-Configured and Dynamic-Configured Beam frame structure. On the basis of NR DRX design, the proposed DRX operation could still be characterized by *ON Duration* (*Active*), *drx-Inactivity Timer* (*Inactive*) and *Opportunity for DRX* (DRX_{sleep}). In active time, the UE with NR DRX turns on its RF module all the time. However, in *Active* and *Inactive*, the UE wake-up pattern follows gNB serving beam pattern. When the UE enters *Opportunity for DRX*, it only performs a short DRX_{check}. If there are any incoming packets, the UE returns to *ON Duration*.

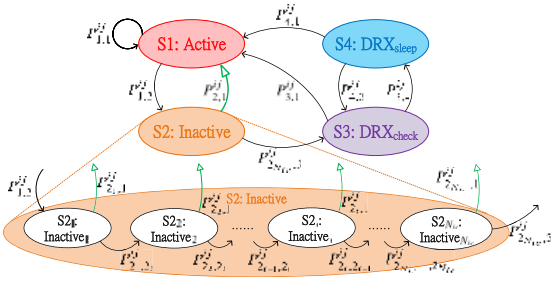


FIGURE 6. The semi-Markov model of UE j on Beam-i with the proposed DRX mechanism, where $P_{2,1}^{ij} = \sum_{m=1}^{N_{ic}} P_{2m,1}^{ij}$.

the configuration of beam patterns, the UEs have the information of the duration in Static-Configured Beam Frame (T_F^s) and Dynamic-Configured Beam Frame (T_F^d). The UEs know when the gNB uses each beam to serve in the Static-Configured Beam Frame, and how to control their RF modules in the Dynamic-Configured Beam Frame. Therefore, the UEs could know when the gNB may serve them and turn on their RF modules accordingly.

The reduction of the unnecessary power consumption: In the Static-Configured Beam Frame, a UE with downlink traffic is served periodically according to the serving beam pattern. When the gNB configures the DRX operation with the awareness of serving beam patterns, the UEs could follow the serving beam patterns to control their RF modules. Thus, the UEs in *Active* and *Inactive* can go to sleep in terms of saving power when the gNB serves other beams in the Static-Configured Beam Frame. In the following sections, we will introduce the semi-Markov Chain for modeling the proposed DRX mechanism (§VI) and analyze the mathematical model (§VII).

VI. THE SEMI-MARKOV CHAIN DRX MODEL

We model the proposed DRX mechanism as the semi-Markov model in Fig. 6. The model includes four states: *Active*, *Inactive*, DRX_{check} and DRX_{sleep}. Considering the serving beam pattern of the gNB, the UE with the proposed DRX mechanism only turns on its RF module when the gNB serves itself. However, the UE with the NR DRX turns on its RF module all the time after receiving a packet arrival. Therefore, the power consumed by the UE in *Active* and *Inactive* is less than that consumed by the UE configured with the 5G NR DRX in *ON Duration*. A DRX cycle consists of one DRX_{check} and one DRX_{sleep}. The details about these states are described as follows:

1) ***Active*** (S1): The UE in *Active* turns on its RF module to monitor PDCCH, transmit data, and receive downlink packets. Unlike the UE behavior in 5G NR networks, the UE in *Active* does not turn on its RF module all the time. Instead, it controls its RF module according to the serving pattern. In Static-Configured Beam Frame, the UE turns on its radio only when the gNB serves it. As for the Dynamic-Configured Beam Frame, the UE will turn on its RF module if the number of transmitted packets exceeds or equals a threshold ($N_{pkt,th}$). At the end of this state, if there are buffered packets, the UE will remain in its current state; otherwise, it will enter *Inactive*.
2) ***Inactive*** ($S2_m$, $\forall m \in \{1, 2, \cdots, N_{ic}\}$): The policy for the UE to control its RF module in *Inactive* is the same as the one in *Active*. As shown in Fig. 6, *Inactive* consists of N_{ic} substates, and the duration of each substate is equal to one serving period (T_F). The *Inactive* is similar to *drx-Inactivity Timer* in the NR DRX, and the number of all substates (N_{ic}) corresponds to the length of the timer. Any transmitted packet in one substate moves the UE to *Active*. If there is no data transmission, the UE will enter the next substate or DRX_{check}.
3) ***DRX_{check}*** (S3): In Fig. 5, a DRX cycle consists of a DRX_{check} and a DRX_{sleep}. The UE in DRX_{check} turns on its RF module for one beam period (T_b) to monitor PDCCH and detect the buffered packets. The length of *Active*, substates in *Inactive*, and the DRX cycles is a multiple of T_F, so the UE enters DRX_{check} at the time t_{begin}. The gNB always serves the UE in the beginning T_b, so the asynchrony problem mentioned in §III can be avoided.
4) ***DRX_{sleep}*** (S4): The UE turns off its RF module for $N_{dc} \cdot T_F - T_b$ in DRX_{sleep} to save power. Every packet arriving in this state is buffered, and the UE enter *Active* to receive buffered data after the DRX cycle. In

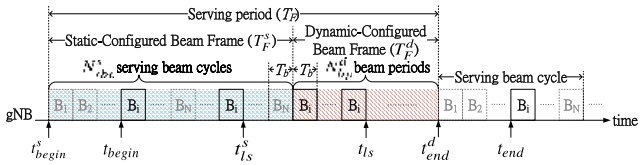


FIGURE 7. The proposed structure of serving beam pattern which consists of a Static-Configured Beam Frame and a Dynamic-Configured Beam Frame, where t_{begin}^s is the beginning of Static-Configured Beam Frame, t_{begin} is the beginning of a state, t_{ls}^s is the last time when the gNB serves the UE in Static-Configured Beam Frame, t_{ls} is the last time when the gNB serves the UE in a state, t_{end}^d is the end of the Dynamic-Configured Beam Frame, and t_{end} is the end of a state.

reality, the UE enters DRX_{check} firstly and then monitors the indicator. In this paper, we use the transition from DRX_{sleep} to *Active* to model the event of the UE receiving the packets buffered in the DRX cycles. If there are no packets sent to the UE, then it will enter DRX_{check}.

VII. MATHEMATICAL FORMULATION

We consider a gNB with N beams and several UEs randomly distributed in the coverage of the gNB. For $i \in \{1, 2, \cdots, N\}$, there are M_i UEs on Beam-i. For $j \in \{1, 2, \cdots, M_i\}$, each packet arrival of UE j on Beam-i follows the Poisson distribution with parameter λ_{ij}, and the random variable of packet inter-arrival time is t_a^{ij}. Considering the extremely high data rate in the mmWave band, we assume that the buffered packets could be transmitted within one beam period. In §IV, the gNB chooses some beams to serve the UEs during the Dynamic-Configured Beam Frame based on the number of packets arriving in the last Static-Configured Beam Frame. The total number of packets sent to all UEs on Beam-i during the interval (t_{begin}^s, t_{ls}^s) in Fig. 7 can be expressed as $n_{pkt}^i = \sum_{j=1}^{M_i} n_{pkt}^{ij}$, where n_{pkt}^{ij} is the random variable of the number of packets sent to UE j on Beam-i in the interval (t_{begin}^s, t_{ls}^s). Also, when there is at least one UE on Beam-i with the packet number no less than threshold $(N_{pkt,th})$, we express the probability as

$$P_i = 1 - \prod_{j=1}^{M_i} \Pr\{n_{pkt}^{ij} < N_{pkt,th}\}.$$

We define the packet number matrix without the packet number of Beam-i as

$$\mathbf{n_{pkt}^{(-i)}} = \begin{bmatrix} n_{pkt}^1 & n_{pkt}^2 & \cdots & n_{pkt}^{i-1} & n_{pkt}^{i+1} & \cdots & n_{pkt}^N \end{bmatrix}^\mathsf{T},$$

and the exceeding probability matrix without the exceeding probability of Beam-i as

$$\mathbf{P^{(-i)}} = diag(\begin{bmatrix} P_1 & P_2 & \cdots & P_{i-1} & P_{i+1} & \cdots & P_N \end{bmatrix}).$$

In addition to Beam-i, if there are another m beams achieving the serving requirements (*i.e.*, existing at least one UE receiving no less than $N_{pkt,th}$ packets) where $m \in \{0, 1, \cdots, N-1\}$, the gNB will select beams to serve in the Dynamic-Configured Beam Frame randomly. The selecting probability is directly proportional to the packet number of each beam. For example, if the threshold is two, the total packet number on Beam-1 is three, and the one on Beam-2 is two. Both of them achieve the serving requirement, and the selecting probabilities are 0.6 and 0.4, respectively. When there are additional m beams satisfying the serving requirements, we express the selecting probability matrix of Beam-i as

$$\mathbf{s_m^{(i)}} = \frac{n_{pkt}^i}{\mathbf{A}(N-1, m)^\mathsf{T} \mathbf{n_{pkt}^{(-i)}}} \qquad \forall m \in \{1, 2, \cdots, N-1\},$$

where every entry is the probability of the gNB using Beam-i to serve under all possible scenarios, and $\mathbf{A}(m, n)$ is a $(0, 1)$-matrix where every column is one of all possible permutations of n ones and $(m-n)$ zeros. $\mathbf{A}(m, n)$ has the following three properties,

1) The size of $\mathbf{A}(m, n)$ is $m \times C_n^m$.
2) $\mathbf{J_{m,1}}^\mathsf{T} \mathbf{A}(m, n) = n \cdot \mathbf{J_{C_n^m,1}}^\mathsf{T}$, where $\mathbf{J_{x,y}}$ is a x-by-y matrix with all its entries equal to one.
3) Let $\mathbf{A}(m, n) = \begin{bmatrix} \mathbf{a_1} & \mathbf{a_2} & \cdots & \mathbf{a_{C_n^m}} \end{bmatrix}$. For $x, y \in \{1, 2, \cdots, C_n^m\}$, $\mathbf{a_x} = \mathbf{a_y}$ if and only if $x = y$.

Besides, we also define two functions to present the derivation in the following. Let $\mathbf{P} \in \mathbb{R}^{n \times n}$, $\mathbf{A} \in \mathbb{R}^{n \times m}$, $\mathbf{a} \in \mathbb{R}^{m \times 1}$, and $m, n \in \mathbb{R}$, one function is

$$f(\mathbf{P}, \mathbf{A}) = \mathbf{J_{n,m}} - \mathbf{A} + \mathbf{P}(2\mathbf{A} - \mathbf{J_{n,m}}),$$

and the other is

$$\Gamma(\mathbf{A}) = \Gamma\left(\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{bmatrix}\right) = \begin{bmatrix} \prod_{i=1}^n a_{i1} & \prod_{i=1}^n a_{i2} & \cdots & \prod_{i=1}^n a_{im} \end{bmatrix}$$

which is used to multiple every entry of each column in a n-by-m matrix, and generate a 1-by-m matrix. We define the vector power operation as

$$\mathbf{a}^{\odot n} = \begin{cases} \mathbf{J_{m,1}}, & \text{if } n = 0 \\ \mathbf{a}, & \text{if } n = 1 \\ diag(\mathbf{a})\mathbf{a}^{\odot(n-1)}, & \text{otherwise} \end{cases}.$$

The probability of Matching occurring in the Dynamic-Configured Beam Frame (*i.e.*, the gNB serves the UE, and the UE turns on its RF module at the same time) is related to the other beams with at least one UE receiving no less than $N_{pkt,th}$ packets. Given the condition that Beam-i achieves the serving requirement, when there are another m beams also achieve the serving requirements, the number of permutations is C_m^{N-1}. With the matrices and functions mentioned above, we can express the probabilities of every permutation happening as a 1-by-C_m^{N-1} matrix

$$\mathbf{p_m^{(-i)}} = \Gamma(f(\mathbf{P^{(-i)}}, \mathbf{A}(N-1, m))) \qquad \forall m \in \{1, 2, \cdots, N-1\}.$$

A. TRANSITION PROBABILITY IN THE SEMI-MARKOV CHAIN MODEL

The mean sojourn time in each state is a random variable instead of a fixed value, so we employ the semi-Markov model in Fig. 6 to analyze the proposed cross-layer DRX mechanism. To solve the semi-Markov Chain and analyze the performance of the proposed DRX, we have to calculate the transition probabilities firstly. For UE j on Beam-i, the transition probability from state Sm to state Sn is denoted as $P^{ij}_{m,n}$, where $m, n \in \mathcal{S}$ and $\mathcal{S} = \{1, 3, 4\} \cup \{2_1, 2_2, \cdots, 2_{N_{ic}}\}$. We will calculate the transition probabilities from the different states as follows:

1) **Transition probability from *Active***: At the end of *Active*, the UE stays in the same state when there are buffered packets; otherwise, the UE enters *Inactive*. The transition probability from *Active* to *Inactive* can be determined with

$$P^{ij}_{1,2_1} = \mathsf{Pr}\{t^{ij}_a \geq \text{interval}\,(t_{ls}, t_{end})\}. \tag{1}$$

The Matching in the Dynamic-Configured Beam Frame impacts on the length of interval (t_{ls}, t_{end}) and $P^{ij}_{1,2_1}$. The expected transition probability from *Active* to *Inactive* is

$$P^{ij}_{1,2_1} = P^{ij}_{(nm)} P^{ij}_{1,2_1(nm)} + \sum_{n=0}^{N^d_{bp}-1} P^{ij,n}_{(m)} P^{ij,n}_{1,2_1(m)} \tag{2a}$$

$$P^{ij}_{1,1} = 1 - P^{ij}_{1,2_1}, \tag{2b}$$

where $P^{ij}_{(nm)}$ is the no Matching probability, $P^{ij}_{(m)}$ is the Matching probability, $P^{ij}_{1,2_1(nm)}$ and $P^{ij}_{1,2_1(m)}$ are the transition probabilities with and without Matching. We will calculate these probabilities in §VII-B.

2) **Transition probability from *Inactive***: If there are no packets sent to the UE in *Inactive*, it will remain in *Inactive* or enter DRX_{check} at the next state. Any packet arrival makes the UE enter *Active* at the end of the state. The length of *Inactive* is T_F. For $m \in \{1, 2, \cdots, N_{ic} - 1\}$,

$$P^{ij}_{2_m,1} = P^{ij}_{2_{N_{ic}},1} = \mathsf{Pr}\{t^{ij}_a < T_F\} = 1 - e^{-\lambda_{ij} T_F} \tag{3a}$$

$$P^{ij}_{2_m,2_{m+1}} = P^{ij}_{2_{N_{ic}},3} = e^{-\lambda_{ij} T_F}. \tag{3b}$$

3) **Transition probability from DRX_{check}**: The packet arrivals let the UE in DRX_{check} move to *Active*. The UE will enter DRX_{sleep} if there are no packets arriving at the UE. The transition probabilities from DRX_{check} to *Active* and DRX_{sleep} are

$$P^{ij}_{3,1} = \mathsf{Pr}\{t^{ij}_a < T_b\} = 1 - e^{-\lambda_{ij} T_b} \tag{4a}$$

$$P^{ij}_{3,4} = 1 - P^{ij}_{3,1} = e^{-\lambda_{ij} T_b}. \tag{4b}$$

4) **Transition probability from DRX_{sleep}**: At the end of DRX_{sleep}, if there are buffered packets, the UE will enter *Active*; otherwise, the UE will enter DRX_{check}. The length of a DRX cycle is $N_{dc} \cdot T_F$, including DRX_{check} and DRX_{sleep}. The probability that there is any packet arrival in DRX_{sleep} is

$$P^{ij}_{4,1} = \mathsf{Pr}\{t^{ij}_a < N_{dc} T_F - T_b\} \tag{5a}$$

$$= 1 - e^{-\lambda_{ij}(N_{dc} T_F - T_b)} \tag{5b}$$

$$P^{ij}_{4,3} = 1 - P^{ij}_{4,1} = e^{-\lambda_{ij}(N_{dc} T_F - T_b)}. \tag{5c}$$

B. THE IMPACT OF MATCHING ON TRANSITION PROBABILITY

In §VI, the state transition at the end of *Active* is based on the existence of buffered packets. In *Active*, every packet before the time t_{ls} in Fig. 5 can be transmitted to the UEs successfully. That is, the packets arriving during the interval (t_{ls}, t_{end}) will be buffered, so the transition probability from *Active* to *Inactive* can be expressed as (1). The time t_{end} is fixed, while whether the Matching occurring in the Dynamic-Configured Beam Frame (*i.e.*, the gNB serves the UE, and the UE turns on its RF module simultaneously) will vary the time t_{ls}. In the following, we will derive $P^{ij}_{1,2_1}$ (2a) by considering Matching and no Matching in *Active* individually.

1) **No Matching in *Active***: If there is no Matching, then UE j on Beam-i would not receive any packet during the Dynamic-Configured Beam Frame. All packets are buffered, and the time t_{ls} is equal to t^s_{ls}. The length of the interval $(t_{ls}, t_{end}) = (t^s_{ls}, t_{end}) = T^d_F + T_b(N-1)$ in this case. Since the packets follow the Poisson distribution, the transition probability from *Active* to *Inactive* is

$$\begin{aligned} P^{ij}_{1,2_1(nm)} &= \mathsf{Pr}\{t^{ij}_a \geq \text{interval}(t_{ls}, t_{end}) | \text{no Matching}\} \\ &= \mathsf{Pr}\{t^{ij}_a \geq T^d_F + T_b(N-1)\} \\ &= e^{-\lambda_{ij}[T^d_F + T_b(N-1)]}. \end{aligned} \tag{6}$$

There are two sufficient conditions that all packet buffered in the Dynamic-Configured Beam Frame. One is UE j on Beam-i turns off its RF module when the packet number is less than $N_{pkt,th}$. The probability of this event is $\mathsf{Pr}\{n^{ij}_{pkt} < N_{pkt,th}\}$. The other is that the gNB does not use Beam-i to serve the UEs in the Dynamic-Configured Beam Frame. Although the number of the packet arrivals for UE j on Beam-i in the interval (t^s_{begin}, t^s_{ls}) is more than or equal to $N_{pkt,th}$, it is still possible that the gNB skips Beam-i and transmits on other beams which also meet the serving requirement in Dynamic-Configured Beam Frame. When the number of the packets transmitted to UE j on Beam-i is no less than $N_{pkt,th}$, the probability of the gNB only serving the UEs on other beams is $\mathsf{Pr}\{n^{ij}_{pkt} \geq N_{pkt,th}\} \cdot \sum_{m=0}^{N-1} \mathbf{p}^{(-\mathbf{i})}_{\mathbf{m}} (1 - \mathbf{s}^{(\mathbf{i})}_{\mathbf{m}})^{\odot N^d_{bp}}$. As a result, the probability of no Matching is

$$\begin{aligned} P^{ij}_{(nm)} &= \mathsf{Pr}\{n^{ij}_{pkt} < N_{pkt,th}\} \\ &\quad + \mathsf{Pr}\{n^{ij}_{pkt} \geq N_{pkt,th}\} \sum_{m=0}^{N-1} \mathbf{p}^{(-\mathbf{i})}_{\mathbf{m}} (1 - \mathbf{s}^{(\mathbf{i})}_{\mathbf{m}})^{\odot N^d_{bp}}. \end{aligned} \tag{7}$$

2) **Matching in *Active***: If there are Matchings in the Dynamic-Configured Beam Frame, the time t_{ls} is the end of the last Matching. We assume that the gNB uses Beam-i to serve UE j at the $(N^d_{bp}-n)^{th}$ beam period in the Dynamic-Configured Beam Frame, and the gNB does not use Beam-i to transmit anymore. The length of interval $(t_{ls}, t_{end}) = (n+i-1)T_b$ which determines the transition probability from *Active* to *Inactive* according to (1). In this case,

$$\begin{aligned} P^{ij,n}_{1,2_1(m)} &= \mathsf{Pr}\{t^{ij}_a \geq \text{interval } (t_{ls}, t_{end}) | \text{Matching}\} \\ &= e^{-\lambda_{ij}[(n+i-1)T_b]}, \end{aligned} \tag{8}$$

when the interval (t_{ls}, t^d_{end}) is nT_b. If the Matching occurs during the Dynamic-Configured Beam Frame, two conditions are satisfied. One is UE j on Beam-i turning on its RF module, and the other is the gNB using Beam-i to transmit data in the Dynamic-Configured Beam Frame. Only when the number of received packet in the interval (t^s_{begin}, t^s_{ls}) exceeds or equals $N_{pkt,th}$ (*i.e.*, achieve the serving requirement) does UE j on Beam-i turn on its RF module, and the probability is $\mathsf{Pr}\{n^{ij}_{pkt} \geq N_{pkt,th}\}$. If the first condition is satisfied, and another m beams also achieve the serving requirement, then the probability that the gNB uses Beam-i to serve the UEs at the $(N^d_{bp}-n)^{th}$ beam period is $\mathbf{p}^{(-\mathbf{i})}_{\mathbf{m}} diag(\mathbf{s}^{(\mathbf{i})}_{\mathbf{m}})(1-\mathbf{s}^{(\mathbf{i})}_{\mathbf{m}})^{\odot n} \quad \forall m \in \{1, 2, \cdots, N-1\}$. The probability of Matching is

$$\begin{aligned} P^{ij,n}_{(m)} = \mathsf{Pr}\{n^{ij}_{pkt} \geq N_{pkt,th}\} \\ \cdot \sum_{m=0}^{N-1} \mathbf{p}^{(-\mathbf{i})}_{\mathbf{m}} diag(\mathbf{s}^{(\mathbf{i})}_{\mathbf{m}})(1-\mathbf{s}^{(\mathbf{i})}_{\mathbf{m}})^{\odot n}. \end{aligned} \tag{9}$$

C. STEADY STATE PROBABILITY

We could derive the solutions of the semi-Markov Chain model, *i.e.*, the steady state probabilities $\Pi^{ij}_m \quad \forall m \in \mathcal{S}$ for UE j on Beam-i, with the transition probabilities. Π^{ij}_m is the stationary distribution over state Sm for UE j on Beam-i in the long run, and we will calculate it in the following Lemma 1.

Lemma 1: The steady state probabilities of UE j on Beam-i are the solutions of the following system of equations

$$\begin{cases} \Pi^{ij}_1 = (\sum_{m\in\mathcal{S}\setminus\{1\}} \Pi^{ij}_m \cdot P^{ij}_{m,1})/(1-P^{ij}_{1,1}) \\ \Pi^{ij}_{2_1} = \Pi^{ij}_1 \cdot P^{ij}_{1,2_1} \\ \Pi^{ij}_{2_m} = \Pi^{ij}_{2_{m-1}} \cdot P^{ij}_{2_{m-1},2_m} \quad \forall m \in \{2, 3, \cdots, N_{ic}\} \\ \Pi^{ij}_3 = (\Pi^{ij}_{2_{N_{ic}}} \cdot P^{ij}_{2_{N_{ic}},3})/(1-P^{ij}_{3,4} \cdot P^{ij}_{4,3}) \\ \Pi^{ij}_4 = \Pi^{ij}_3 \cdot P^{ij}_{3,4} \\ \sum_{m\in\mathcal{S}} \Pi^{ij}_m = 1. \end{cases} \tag{10}$$

Proof: The equation of steady state probability for UE j on beam i in each state is

$$\Pi^{ij}_m = \sum_{n\in\mathcal{S}} \Pi^{ij}_n \cdot P^{ij}_{n,m} \quad \forall m \in \mathcal{S}. \tag{11}$$

Thus, the steady state probability of *Active* can be expressed as

$$\Pi^{ij}_1 = \frac{\sum_{m\in\mathcal{S}\setminus\{1\}} \Pi^{ij}_m \cdot P^{ij}_{m,1}}{1-P^{ij}_{1,1}}. \tag{12}$$

Fig. 6 shows that the steady state probabilities of other states are

$$\Pi^{ij}_{2_1} = \Pi^{ij}_1 \cdot P^{ij}_{1,2_1} \tag{13a}$$

$$\Pi^{ij}_{2_m} = \Pi^{ij}_{2_{m-1}} \cdot P^{ij}_{2_{m-1},2_m} \quad \forall m \in \{2, 3, \cdots, N_{ic}\} \tag{13b}$$

$$\Pi^{ij}_3 = \Pi^{ij}_{2_{N_{ic}}} \cdot P^{ij}_{2_{N_{ic}},3} + \Pi^{ij}_4 \cdot P^{ij}_{4,3} \tag{13c}$$

$$\Pi^{ij}_4 = \Pi^{ij}_3 \cdot P^{ij}_{3,4}. \tag{13d}$$

With (13d), the steady state probability of DRX_{check} (13c) can be expressed as

$$\Pi^{ij}_3 = \Pi^{ij}_{2_{N_{ic}}} P^{ij}_{2_{N_{ic}},3} + \Pi^{ij}_3 P^{ij}_{3,4} P^{ij}_{4,3} = \frac{\Pi^{ij}_{2_{N_{ic}}} P^{ij}_{2_{N_{ic}},3}}{1-P^{ij}_{3,4}P^{ij}_{4,3}}. \tag{14}$$

The last equation is the constraint that the summation of all steady state probabilities is equal to one. □

D. SLEEP RATIO

We can use sleep ratio (R^{ij}_{sleep}) and packet delay (D^{ij}_{pkt}) to evaluate DRX mechanism. For UE j on Beam-i, R^{ij}_{sleep} is the proportion of accumulation of *Opportunity for DRX* to the total observed time. Higher R^{ij}_{sleep} represents lower power consumption. To derive $\mathsf{E}\{R^{ij}_{sleep}\}$, we have to get the steady state probability ($\Pi^{ij}_m \quad \forall m \in \mathcal{S}$), the expected holding time ($\mathsf{E}\{T_{h,m}\} \quad \forall m \in \mathcal{S}$), and sleeping time in *Active* and *Inactive* (T^{ij}_{off}). Lemma 1 yields Π^{ij}_m, and the derivations of $\mathsf{E}\{T_{h,m}\}$ and T^{ij}_{off} will be showed as follows:

Lemma 2: The expected holding time for each state of each UE is

$$\begin{aligned} \mathsf{E}\{T_{h,1}\} &= T_F \\ \mathsf{E}\{T_{h,2_m}\} &= T_F \quad \forall m \in \{1, 2, \cdots, N_{ic}\} \\ \mathsf{E}\{T_{h,3}\} &= T_b \\ \mathsf{E}\{T_{h,4}\} &= T_F \cdot N_{dc} - T_b. \end{aligned} \tag{15}$$

Proof: No matter which state the UEs stay currently, they will enter *Active* or other states at the end of their current states. Fig. 5 shows that the holding times in *Active* and *Inactive* are T_F, and the length of DRX cycle is $T_F \cdot N_{dc}$. The UEs will turn on their RF module for one beam period at the beginning of each DRX cycle to check whether there is any buffered packet or not, so $\mathsf{E}\{T_{h,3}\}$ is T_b and $\mathsf{E}\{T_{h,4}\}$ is $T_F \cdot N_{dc} - T_b$. □

Lemma 3: The expected sleeping time of UE j on Beam-i in Active and Inactive is

$$\mathsf{E}\{T^{ij}_{off}\} = T^s_F - N^s_{sbc} T_b + \mathsf{Pr}\{n^{ij}_{pkt} < N_{pkt,th}\} T^d_F. \tag{16}$$

Proof: *Active* and *Inactive* consist of one Static-Configured and one Dynamic-Configured Beam Frame. In the Static-Configured Beam Frame, the UEs turn on their

RF module only when the gNB serves them. The gNB uses Beam-i to serve UE j for one beam period per one time and N^s_{sbc} times in one Static-Configured Beam Frame, so the sleeping time in Static-Configured Beam Frame is $T^s_F - N^s_{sbc} \cdot T_b$. The UEs will turn off their RF modules in Dynamic-Configured Beam Frame if the number of received packets is less than $N_{pkt,th}$ in the interval (t^s_{begin}, t^s_{ls}). The average sleeping time of UE j on Beam-i in Dynamic-Configured Beam Frame is $\mathsf{Pr}\{n^{ij}_{pkt} < N_{pkt,th}\} \cdot T^d_F$. □

Theorem 1: The expected sleep ratio of UE j on Beam-i is

$$\mathsf{E}\{R^{ij}_{sleep}\} = \frac{(\Pi^{ij}_1 + \sum_{i=1}^{N_{ic}} \Pi^{ij}_{2_i})\mathsf{E}\{T^{ij}_{off}\} + \Pi^{ij}_4 \mathsf{E}\{T_{h,4}\}}{\sum_{m\in\mathcal{S}} \Pi^{ij}_m \cdot \mathsf{E}\{T_{h,m}\}}. \quad (17)$$

Proof: The sleep ratio is the proportion of the time that the UE turns off its RF module. From (16), we can get the sleeping time of UE j on Beam-i in *Active* and *Inactive* is T^{ij}_{off}. The UEs turn on their RF modules in DRX_{check} and turn off their RF module in DRX_{sleep}. Considering the sleeping time in every state and the steady state probability from (10), we can express the expected sleep ratio as (17). □

E. PACKET DELAY

If a downlink packet arrives during the sleeping time of the UE, it will be buffered, and the delay (D^{ij}_{pkt}) increases. D^{ij}_{pkt} is the period from the time when the gNB receives a packet to the time when it uses Beam-i to transmit the packet to UE j. In the following, we will calculate the expected delay in different states, and then derive $\mathsf{E}\{D^{ij}_{pkt}\}$ with the expected delay and steady state probability.

Lemma 4: The average delay for a packet arriving at the UEs in DRX_{sleep} is

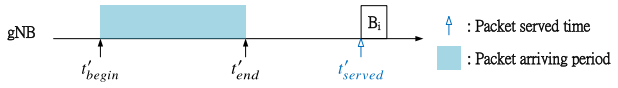
$$d_{sleep} = \frac{1}{N_{dc}T_F - T_b} \int_{T_b}^{N_{dc}T_F} (N_{dc}T_F - t)\, dt. \quad (18)$$

Proof: According to [23], if the UEs receive all buffered packets at the time t'_{served}, the average delay for a packet in the interval (t'_{begin}, t'_{end}) in Fig. 8(a) will be

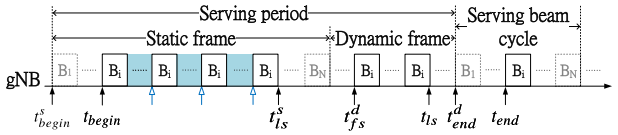
$$\int_{t'_{begin}}^{t'_{end}} (t'_{served} - t)\mathsf{Pr}\{t_a = t\}\, dt = \frac{1}{t'_{end} - t'_{begin}} \int_{t'_{begin}}^{t'_{end}} (t'_{served} - t)\, dt \quad (19)$$

where t_a is the packet arriving time. In DRX_{sleep}, $t'_{begin} = T_b$ and $t'_{end} = t'_{served} = N_{dc} \cdot T_F$. The average delay in DRX cycles is (18). □

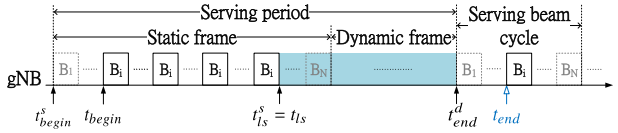
To get $\mathsf{E}\{D^{ij}_{pkt}\}$, we need to derive the delay in *Active*, *Inactive*, and DRX_{sleep}. The delay in *Active* equals the one in *Inactive*, and would be influenced by Matchings (*i.e.*, the gNB serves the UE, and it turns on its radio at the same time). Fig. 8 demonstrates the delay under different cases in *Active* and *Inactive*. The period of an *Active* or an *Inactive* consists of four intervals, including (t_{begin}, t^s_{ls}), the duration from t^s_{ls} to the end of Static-Configured Beam Frame, the duration from the end of Static-Configured Beam Frame to t^d_{end}, and (t^d_{end}, t_{end}).



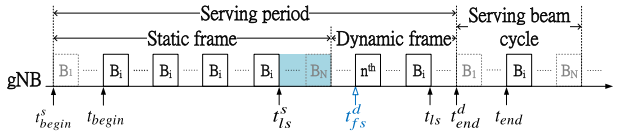
(a) The delay model for the packets arriving in the blue block and buffered, and the hollow blue arrow represents the time when the UEs can receive the buffered packets.



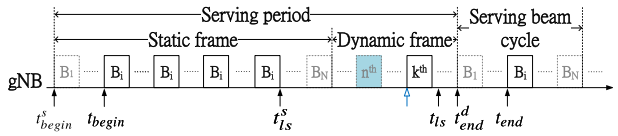
(b) The delay $d_{(s)}$ for the packets arriving before the time t^s_{ls}.



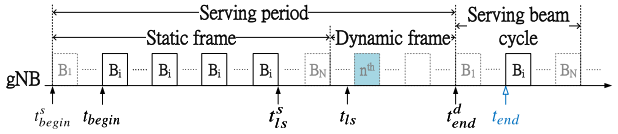
(c) The delay $d^{ij}_{(nm)}$ for the packets arriving in the interval (t^s_{ls}, t^d_{end}) if there is no matching in the Dynamic Frame.



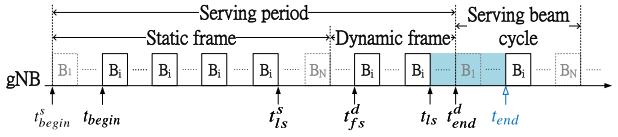
(d) The delay $d^{ij}_{(m)}$ for the packets arriving between the time t^s_{ls} and the beginning of the Dynamic Frame, and the first matching occurs at the n^{th} beam period.



(e) The delay $d^{ij}_{(d,m)}$ for the packets arriving in the n^{th} beam period of the Dynamic Frame and the matching occurs at the k^{th} beam period.



(f) The delay $d^{ij}_{(d,nm)}$ for the packets arriving in the n^{th} beam period if there is no matching occurring after the n^{th} beam period in the Dynamic Frame.



(g) The delay d^{ij}_{rest} for the packets arriving in the interval (t^d_{end}, t_{end}).

FIGURE 8. The delay model in *Active* and *Inactive*. We use "Static Frame" and "Dynamic Frame" to presented "Static-Configured Beam Frame" and "Dynamic-Configured Beam Frame".

With the consideration of Matching and the packet arriving period, the expected delay in *Active* or *Inactive* is expressed as

$$d^{ij}_{act} = \frac{d_{(s)} + d^{ij}_{(nm)} + d^{ij}_{(m)} + d^{ij}_{(d,m)} + d^{ij}_{(d,nm)} + d^{ij}_{rest}}{T_F}. \quad (20)$$

In the following, we will derive the delay based on the cases in Fig. 8.

1) $d_{(s)}$: Fig. 8(b) shows the packets for UE j on Beam-i arriving in the serving beam cycles are buffered when

the gNB transmit on other beams. There are $N^s_{sbc}-1$ serving beam cycles in the interval (t_{begin}, t^s_{ls}), and the delay in each serving beam cycle is $\int_{T_b}^{N\cdot T_b}(N\cdot T_b - t)\,dt$. The delay in the interval (t_{begin}, t^s_{ls}) is

$$d_{(s)} = (N^s_{sbc}-1)\int_{T_b}^{N\cdot T_b}(N\cdot T_b - t)\,dt. \tag{21}$$

2) $d^{ij}_{(nm)}$: For the packet arrivals in the interval (t^s_{ls}, t^d_{end}), we have to consider the Matchings in the Dynamic-Configured Beam Frame when we calculate the delay. If there is no Matching in the Dynamic-Configured Beam Frame, $t^s_{ls} = t_{ls}$ and all packets arriving in the interval (t^s_{ls}, t^d_{end}) will be buffered until the time t_{end}. There is no Matching occurring only when the received packet number in the interval (t^s_{begin}, t^s_{ls}) is less than $N_{pkt,th}$, or the gNB transmits on other beams in the Dynamic-Configured Beam Frame. With the consideration of the probability of no Matching in the Dynamic-Configured Beam Frame, the delay in Fig. 8(c) is derived as

$$\begin{aligned} d^{ij}_{(nm)} = {} & \left\{\mathsf{Pr}\{n^{ij}_{pkt} < N_{pkt,th}\} + \mathsf{Pr}\{n^{ij}_{pkt} \geq N_{pkt,th}\}\right. \\ & \left.\cdot\left[\sum_{m=1}^{N-1}\mathbf{p}^{(-\mathbf{i})}_{\mathbf{m}}(1-\mathbf{s}^{(\mathbf{i})}_{\mathbf{m}})^{\odot N^d_{bp}}\right]\right\} \\ & \cdot\int_{iT_b}^{N\cdot T_b+T^d_F}[N\cdot T_b + T^d_F + (i-1)T_b - t]\,dt. \end{aligned} \tag{22}$$

3) $d^{ij}_{(m)}$: In the Static-Configured Beam Frame, if the packets arrive after the time t^s_{ls}, they will be buffered. Fig. 8(d) shows that there are Matchings occurring, so UE j on Beam-i receives the buffered packets in the following Dynamic-Configured Beam Frame. In this case, the number of received packets in the interval (t^s_{begin}, t^s_{ls}) of UE j on Beam-i is no less than $N_{pkt,th}$, and the gNB uses Beam-i to serve UE j in the n^{th} beam period of the Dynamic-Configured Beam Frame. When both Beam-i and another m beams achieve the serving requirements, the probability that the gNB uses Beam-i to serve UE j at the n^{th} beam period is $\mathsf{Pr}\{n^{ij}_{pkt} \geq N_{pkt,th}\cdot \mathbf{p}^{(-\mathbf{i})}_{\mathbf{m}}\mathbf{s}^{(\mathbf{i})}_{\mathbf{m}}(1-\mathbf{s}^{(\mathbf{i})}_{\mathbf{m}})^{\odot(n-1)}$. Thus, the delay in Fig. 8(d) is expressed as

$$\begin{aligned} d^{ij}_{(m)} = {} & \mathsf{Pr}\{n^{ij}_{pkt} \geq N_{pkt,th}\} \\ & \sum_{m=0}^{N-1}\sum_{n=1}^{N^d_{bp}}\mathbf{p}^{(-\mathbf{i})}_{\mathbf{m}}\mathbf{s}^{(\mathbf{i})}_{\mathbf{m}}(1-\mathbf{s}^{(\mathbf{i})}_{\mathbf{m}})^{\odot(n-1)} \\ & \cdot\int_{iT_b}^{N\cdot T_b}[N\cdot T_b + (n-1)T_b - t]\,dt. \end{aligned} \tag{23, 24}$$

4) $d^{ij}_{(d,m)}$: Fig. 8(e) illustrates the packets arrive in the n^{th} beam period of the Dynamic-Configured Beam Frame, and they are buffered until the gNB uses Beam-i to transmit data in the k^{th} beam period. The probability of this case is $\mathsf{Pr}\{n^{ij}_{pkt} \geq N_{pkt,th}\cdot \mathbf{p}^{(-\mathbf{i})}_{\mathbf{m}}\mathbf{s}^{(\mathbf{i})}_{\mathbf{m}}(1-\mathbf{s}^{(\mathbf{i})}_{\mathbf{m}})^{\odot(k-n)}$, where there are both UE j on Beam-i and another m beams satisfy the serving requirement. The delay in this case is

$$\begin{aligned} d^{ij}_{(d,m)} = {} & \mathsf{Pr}\{n^{ij}_{pkt} \geq N_{th}\}\sum_{m=0}^{N-1}\sum_{n=1}^{N^d_{bp}-1}\sum_{k=n+1}^{N^d_{bp}}\mathbf{p}^{(-\mathbf{i})}_{\mathbf{m}} \\ & \mathbf{s}^{(\mathbf{i})}_{\mathbf{m}}(1-\mathbf{s}^{(\mathbf{i})}_{\mathbf{m}})^{\odot(k-n)}\int_{(n-1)T_b}^{nT_b}(kT_b - t)\,dt. \end{aligned} \tag{25}$$

5) $d^{ij}_{(d,nm)}$: In Fig. 8(f), all Matchings happen before the n^{th} beam period, so the packets for UE j on Beam-i in the n^{th} beam period in the Dynamic-Configured Beam Frame are buffered until the time t_{end}. Since the gNB only uses other beams to transmit, so there must be another $m \neq 0$ beams achieve the serving requirement. The delay in Fig. 8(f) is expressed as

$$\begin{aligned} d^{ij}_{(d,nm)} = {} & \mathsf{Pr}\{n^{ij}_{pkt} \geq N_{th}\} \\ & \cdot\sum_{m=0}^{N-1}\sum_{n=2}^{N^d_{bp}}\mathbf{p}^{(-\mathbf{i})}_{\mathbf{m}}[1-(1-\mathbf{s}^{(\mathbf{i})}_{\mathbf{m}})^{\odot(n-1)}] \\ & \cdot\int_{(n-1)T_b}^{nT_b}[(N^d_{bp}-n+i)T_b - t]\,dt. \end{aligned} \tag{26}$$

6) d^{ij}_{rest}: The packets arriving in the interval (t^d_{end}, t_{end}) are buffered until the end of the state, so the delay in Fig. 8(g) is derived as

$$d^{ij}_{rest} = \int_0^{(i-1)T_b}[(i-1)T_b - t]\,dt. \tag{27}$$

Theorem 2: The expected delay for one packet of UE j on beam i is

$$\begin{aligned} & \mathsf{E}\{D^{ij}_{pkt}\} \\ & = \frac{\sum_{m\in\mathcal{S}\setminus\{3,4\}}\Pi^{ij}_m\mathsf{E}\{T_{h,m}\}d^{ij}_{act} + \Pi^{ij}_4\mathsf{E}\{T_{h,4}\}d_{sleep}}{\sum_{m\in\mathcal{S}}\Pi^{ij}_m\cdot\mathsf{E}\{T_{h,m}\}}. \end{aligned} \tag{28}$$

Proof: When the UE turns off its RF module, the arriving packets are buffered, and the delay increases. The packet delay consists of the delay in each state of the proposed DRX mechanism. The overall average packet delay is the weighted summation of the average delay in each state, and the weights are the probabilities of a packet arriving in the different states. Fig. 6 shows that the packet delay only occurs in *Active*, *Inactive*, and DRX_{sleep}. The probability of a packet arriving in state Sm is

$$\frac{\Pi^{ij}_m\cdot\mathsf{E}\{T_{h,m}\}}{\sum_{n\in\mathcal{S}}\Pi^{ij}_n\cdot\mathsf{E}\{T_{h,n}\}} \quad \forall m\in\mathcal{S}. \tag{29}$$

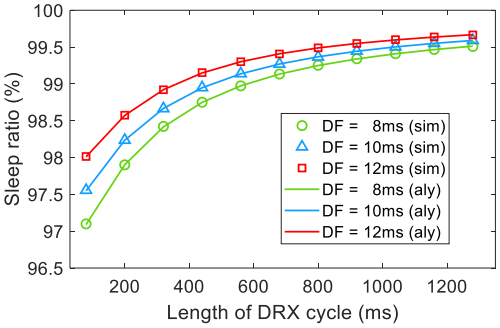
We can get the packet delay in *Active*, *Inactive*, and DRX_{sleep} via Lemma 4 and (20). Thus, the expected delay for one packet is (28). □

VIII. NUMERICAL ANALYSIS AND SIMULATION RESULTS

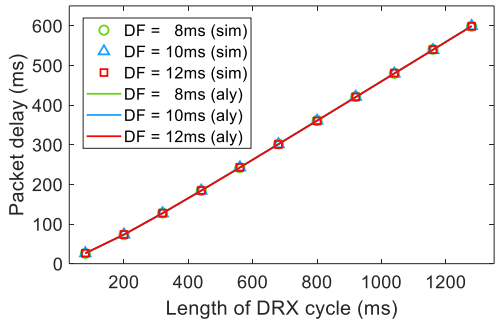
In this section, we verify the analytical model with the simulation experiments conducted on MATLAB. In the downlink

TABLE 1. Default parameter setting.

Symbols	The meaning of the symbol	Default value
N	Total beam number of one gNB	8
M_i	Total UE number on beam i $\forall i \in \{1, 2, \cdots, N\}$	2
λ_{ij}	Inter-arrival rate of UE j on beam i	$0.01\ ms^{-1}$
$N_{pkt,th}$	Threshold of packet number	3
N_{ic}	The total number of substates in *Inactive*	2
	The length of DRX cycle	$960\ ms$
T_F	The length of one serving period	$20\ ms$
T_b	The length of one beam period	$0.125\ ms$



(a) Sleep ratio with different N_{dc}, N^s_{sbc}, and N^d_{bp}.



(b) Packet delay with different N_{dc}, N^s_{sbc}, and N^d_{bp}.

FIGURE 9. The simulation and analytical results for the proposed DRX, where SF is 20−DF *ms*.

scenario we consider, there is one gNB with multiple beams, and the range of beam number is from 4 beams to 22 beams. Besides, there are two UEs on each beam. For the serving beam pattern, we set a beam period (T_b) as 1 slot. We consider the fourth numerology in the 3GPP NR frame structure, so the slot length is 0.125 *ms* [24]. For the data transmission rate in mmWave band, we follow the assumption in § VII, and the transmission time is short enough to be ignored. While fixing the length of the serving period (T_F) as 20 *ms*, we change the number of the serving beam cycle in Static-Configured Beam Frame (N^s_{sbc}) and the number of beam period in Dynamic-Configured Beam Frame (N^d_{bp}) to evaluate the performance of the proposed DRX mechanism. The packet arrival process is Poisson distribution with $\lambda^{ij} = 0.01\ ms^{-1}$ or $\lambda^{ij} = 0.03\ ms^{-1}$ for each UE, and we utilize the threshold of the packet number ($N_{pkt,th}$) = 3. As for the DRX configuration, the maximal number of *Inactive* (N_{ic}) is from 1 to 5, and the number of serving period in DRX cycle (N_{dc}) varies from 4 to 64. Therefore, the maximal length of *Inactive* is from 20 *ms* to 120 *ms*, and the length of DRX cycle is from 80 *ms* to 1280 *ms*. In the following sections, the unmentioned values of parameters are set to the default parameter setting listed in Table 1. In addition, we refer the length of Static-Configured Beam Frame as "SF" and that of Dynamic-Configured Beam Frame as "DF".

A. ANALYTICAL MODEL VERIFICATION AND EFFECT OF DRX CYCLE

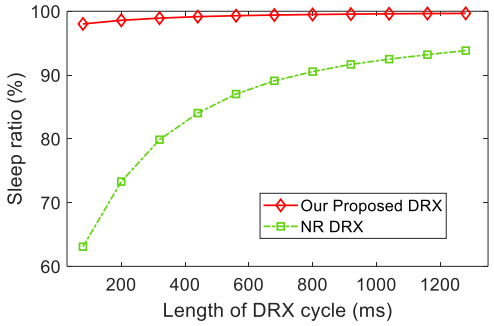
We change the length of DRX cycle and the DF (T^d_F) to examine the effects of these parameters. The serving period ($T_F = T^s_F + T^d_F$) is set to 20 *ms*, so the SF (T^s_F) will change when we use the different DF (T^d_F). Then, we validate the analytical model with simulation results. In Fig. 9(a), the sleep ratio increases, as the length of DRX cycle increases. When the length of DRX_{check} is fixed, longer DRX cycles implies a longer length of DRX_{sleep}. Fig. 9(b) shows that the packets arriving in the longer DRX cycles are buffered, and yield more serious delay than those under the configuration with shorter DRX cycles. That is to say, the longer the DRX cycle length is, the more efficient the power usage is, and the more serious the packet delay is. The tradeoff between power efficiency and packet delivery delay exists with different DRX configurations.

B. COMPARISON WITH OTHER DRX MECHANISMS

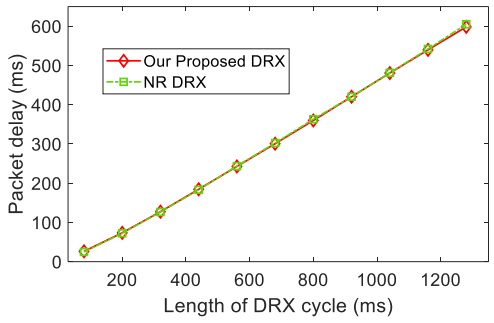
We compare the proposed DRX design with the 5G NR DRX mechanism under the different parameter settings, including the length of DRX cycle, the maximal duration in *Inactive*, the traffic density, and the beam number of the gNB. In §II-A, the UEs with basic DRX operation control their RF modules based on the packet arrivals. When there are packets buffered in the gNB, and the destination UE turns on its RF module, the gNB will use the correct beam to schedule data. Without considering the beam patterns in the 5G networks, the NR DRX may cause the inefficiency of power consumption on UE side. The length of *drx-Inactivity Timer* in the NR DRX corresponds to the duration of one *Active* (T_F) and the maximal period in *Inactive* ($N_{ic} \cdot T_F$) in the proposed DRX mechanism. In this paper, both of them are referred to "the length of Inactivity Timer". In the following, we set the length of *drx-Inactivity Timer* in the NR DRX as $N_{ic} \cdot T_F + T_F$ *ms*.

1) EFFECT OF DRX CYCLE

For the proposed DRX mechanism, we configure the number of the serving beam cycle in Static-Configured Beam Frame (N^s_{sbc}) as 8 and the number of the beam period in Dynamic-Configured Beam Frame (N^d_{bp}) as 96, so the serving period (T_F) is 20 *ms*. Since the maximal number of substates in *Inactive* (N_{ic}) is 2, the maximal length in *Inactive* is 40 *ms*. The length of *drx-Inactivity Timer* in the NR DRX mechanism is set to 60 *ms*. In Fig. 10, the active UE with NR DRX turns on its RF module all the time, so its sleep ratio is smaller than the proposed DRX mechanisms. With the consideration of serving beam pattern, the sleep ratio in our proposed cross-layer

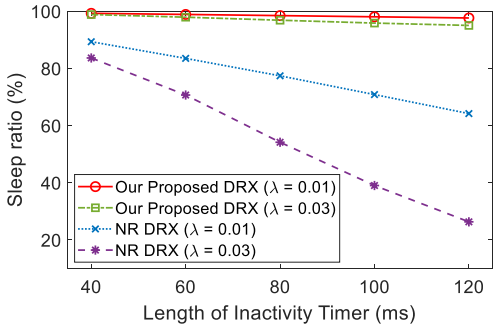


(a) Sleep ratio with different N_{dc}.

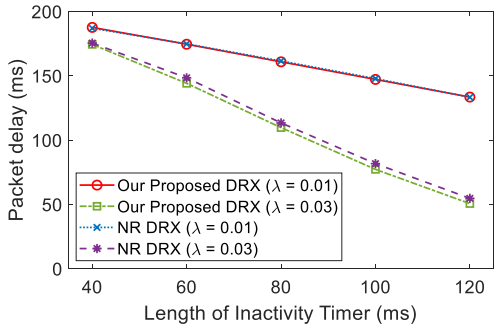


(b) Packet delay with different N_{dc}.

FIGURE 10. The simulation results for the proposed DRX and the 5G NR DRX mechanisms, where the length of Inactivity Timer is 60 *ms*.



(a) Sleep ratio with different N_{ic} and packet arrival rates.



(b) Packet delay with different N_{ic} and packet arrival rates.

FIGURE 11. The simulation results for the proposed DRX and 5G NR DRX mechanism, where λ presents the packet arrival rate of each UE.

DRX mechanism has 41.6% improvement when the DRX cycle is 80 *ms*. The overall performance of our proposed DRX can achieve up to 99.6%. Fig. 10(b) demonstrates that the packet delay in NR DRX is almost the same as the delay in our proposed scheme. As a result, our proposed DRX mechanism can reduce power consumption without producing a larger delay. In the 5G mmWave networks, taking the directional characteristic of the beam pattern into account is essential for the DRX design; otherwise, the unnecessary power consumption mentioned in §III will reduce the efficiency.

2) EFFECT OF INACTIVITY TIMER

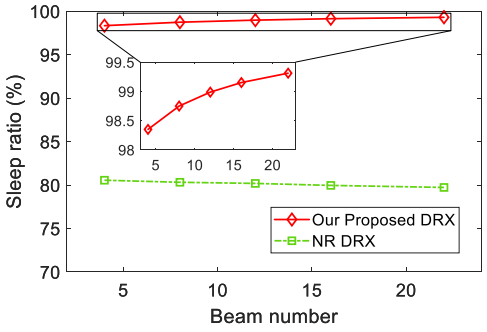
In addition to the length of the DRX cycle, the length of Inactivity Timer plays a critical role in the performance of the DRX mechanism. In Fig. 11, we configure the number of the serving beam cycle in Static-Configured Beam Frame (N^s_{sbc}) as 8 and the number of the beam period in Dynamic-Configured Beam Frame (N^d_{bp}) as 96, so the serving period ($T_F = T^s_F + T^d_F$) is 20 *ms*. The length of DRX cycle is set to 960 *ms* and the length of Inactivity Timer varies from 40 *ms* to 120 *ms*. Fig. 11 depicts that both the sleep ratio and the packet delay of the proposed DRX mechanism decrease when N_{ic}, i.e., the maximal number of substates in *Inactive*, increases. The UEs in *Active* and *Inactive* consume more power than that in DRX_{check} and DRX_{sleep}, while the packet delay in *Active* and *Inactive* is shorter than that in DRX_{check} and DRX_{sleep}. If N_{ic} rises, the probability of entering DRX_{check} and DRX_{sleep} will decline. Therefore, the UEs with higher N_{ic} have higher probability staying in *Active* and *Inactive*, and they cause the lower sleep ratio and lower delay. Compared with the NR DRX, the proposed DRX has more then 11.1% improvement in the energy efficiency without the larger packet delay when the packet arrival rate is 0.01 ms^{-1}.

3) EFFECT OF TRAFFIC LOAD

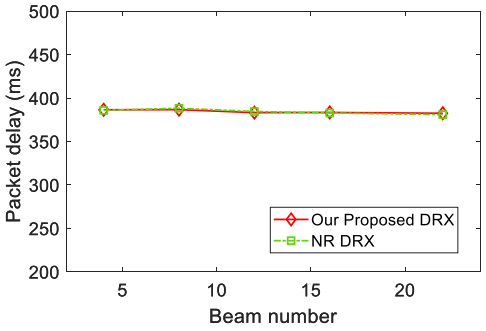
The different curves in Fig. 11 present the performance of the proposed DRX mechanism and the NR DRX under different packet arrival rates, including 0.01 ms^{-1} and 0.03 ms^{-1}. More packets arrive in a fixed period under the heavier traffic load. A packet arriving in *Active* and *Inactive* makes the UEs extend their wake-up period, and one arriving in DRX cycles triggers the UEs to enter *Active*. The frequent packet arrivals make the steady state probabilities of *Active* and *Inactive* high. Thus, both the sleep ratio and the packet delay are lower than those under lighter traffic load in Fig. 11. With the proposed DRX, the UEs could turn off their RF module to save more power when the gNB is serving other UEs. Under the different traffic loads, the UEs with the proposed DRX mechanism reduce more power consumption than that in the NR DRX.

4) EFFECT OF BEAM NUMBER

In Fig. 12, we evaluate the performance of the proposed DRX and the NR DRX mechanism under the number of beams of the gNB from 4 to 22. The number of the serving beam cycles in Static-Configured Beam Frame (N^s_{sbc}) and the beam period in Dynamic-Configured Beam Frame (N^d_{bp}) are configured as 4 and 32. Since the SF is related to the beam number, the serving period (T_F) varies from 6 *ms* to 15 *ms* with different beam numbers. For the DRX configuration, we set the length

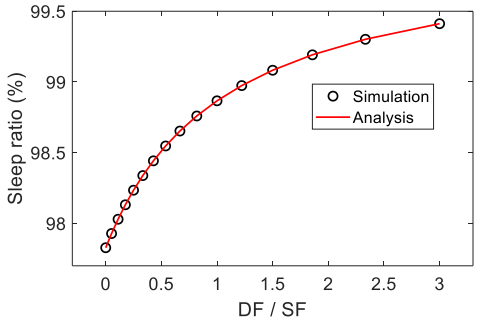


(a) Sleep ratio with different beam numbers.

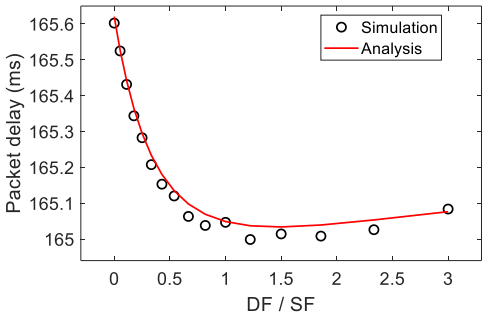


(b) Packet delay with different beam numbers.

FIGURE 12. The simulation results for the proposed DRX and the 5G NR DRX mechanism, where the length of DRX cycle is 960 *ms* and the length of Inactivity Timer is 120 *ms*.



(a) Sleep ratio with various N^d_{bp} and N^s_{sbc}.



(b) Packet delay with various N^d_{bp} and N^s_{sbc}.

FIGURE 13. The simulation and analytical results for the proposed DRX mechanisms, where SF is the length of Static-Configured Beam Frame and DF is the length of Dynamic-Configured Beam Frame, and the length of DRX cycle is 400 *ms*.

of DRX cycle and the length of Inactivity Timer to 960 *ms* and 120 *ms*. Fig. 12(a) depicts the sleep ratio in the proposed DRX increases with the increasing of the beam number. Since N^s_{sbc} is fixed and the UEs sleep when the gNB uses other beams to serve in Static-Configured Beam Frame, the UEs could save more power with the larger beam number. For the NR DRX, only the DRX configuration and the packet arrival rate affect the performance, so the sleep ratio maintains the same value with different beam numbers. Compared with the NR DRX, the proposed DRX could enhance up to 29.7% in the sleep ratio. Also, the sleep ratio in the proposed DRX is higher with the larger beam number. It is essential to design a DRX mechanism with the awareness of beam patterns in mmWave communications. In Fig. 12(b), the packet delay is insensitive to the beam number for both the proposed DRX and the NR DRX mechanism.

C. OPTIMAL DYNAMIC-CONFIGURED BEAM FRAME SETTING

The different curves shown in Fig. 9 serve as the different DF. These curves represent that the performance in the longest DF has the highest sleep ratio with the same packet delay. The DF likely affects the performance of our proposed DRX. The flexible resource in the Dynamic-Configured Beam Frame can improve the QoS for UEs in urgent of data transmission. We set the length of DRX cycle to 400 *ms* and change the ratio of the DF to the SF (T^d_F/T^s_F) to evaluate the performance.

Fig. 13 demonstrates the performance under different ratios of the DF to the SF. While providing the opportunity of transmission for the UEs receiving more packets in the last Static-Configured Beam Frame, the longer DF will cause more serious delay for those whom the gNB does not serve in the Dynamic-Configured Beam Frame. With the increasing DF, the sleep ratio becomes higher, and an optimal packet delay exists. Thus, Fig. 13(a) shows that the sleep ratio increases as the ratio of DF to SF increases. In Fig. 13(b), setting the ratio of DF to SF to 1.5, which means DF is 12 *ms* and SF is 8 *ms*, can minimize the packet delay. When the DF is less than 12 *ms*, the longer Dynamic-Configured Beam Frame can reduce the overall packet delay effectively. When DF is larger than 12 *ms*, the increasing delay of those whom the gNB does not serve is larger than the reduction of packet delay of those whom the gNB serves in Dynamic-Configured Beam Frame. Therefore, the delay increases with the longer Dynamic-Configured Beam Frame. The length of Dynamic-Configured Beam Frame has a significant impact on the performance of our proposed cross-layer DRX.

IX. CONCLUSIONS

In order to satisfy the high requirement of the network performance in 5G wireless communications, the beamforming technology is widely applied and brings the directionality issue to the wireless communication links. Moreover, massive antenna arrays are used in mmWave bands to form a more narrow beam to overcome the severe path loss. However, such link directionality makes the existing DRX inefficient and wastes precious energy. Because of the directionality, a UE

is likely to be "out-of-beam-coverage" in its *ON Duration* when the gNB is serving the UEs on other beams.

In this work, we proposed a dynamic beam-aware DRX framework considering the strategies of gNB serving beam patterns. We use a semi-Markov chain to model the behavior of UEs and validate the model by simulations. The results show that the asynchrony between serving beam patterns and UE wake-up timing significantly impairs the performance of NR DRX. The proposed cross-layer DRX provides up to 41.6% improvement in the sleep ratio without a more considerable packet delay.

REFERENCES

[1] *IMT Vision–Framework and Overall Objectives of the Future Development of IMT for 2020 and Beyond*, document Recommendation ITU 0-2083, 2015.

[2] A. Ghosh, T. A. Thomas, M. C. Cudak, R. Ratasuk, P. Moorut, F. W. Vook, T. S. Rappaport, G. R. MacCartney, S. Sun, and S. Nie, "Millimeter-wave enhanced local area systems: A high-data-rate approach for future wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1152–1163, Jun. 2014.

[3] S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter-wave cellular wireless networks: Potentials and challenges," *Proc. IEEE*, vol. 102, no. 3, pp. 366–385, Mar. 2014.

[4] S. Sur, I. Pefkianakis, X. Zhang, and K.-H. Kim, "Towards scalable and ubiquitous millimeter-wave wireless networks," in *Proc. 24th Annu. Int. Conf. Mobile Comput. Netw.*, 2018, pp. 257–271.

[5] A. Thornburg, T. Bai, and R. W. Heath, Jr., "Interference statistics in a random mmWave ad hoc network," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2015, pp. 2904–2908.

[6] H. Shokri-Ghadikolaei, C. Fischione, G. Fodor, P. Popovski, and M. Zorzi, "Millimeter wave cellular networks: A MAC layer perspective," *IEEE Trans. Commun.*, vol. 63, no. 10, pp. 3437–3458, Oct. 2015.

[7] M. Giordani, M. Polese, A. Roy, D. Castor, and M. Zorzi, "A tutorial on beam management for 3GPP NR at mmWave frequencies," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 1, pp. 173–196, 1st Quart., 2019.

[8] R. W. Heath, Jr., "Millimeter wave: The future of commercial wireless systems," in *Proc. IEEE Compound Semiconductor Integr. Circuit Symp. (CSICS)*, Oct. 2016, pp. 1–4.

[9] F. Sohrabi and W. Yu, "Hybrid digital and analog beamforming design for large-scale antenna arrays," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 501–513, Apr. 2016.

[10] *Study on UE Power Saving in NR*, document TR 38.840, 3rd Generation Partnership Project (3GPP), 2019.

[11] C.-H. Ho, A. Huang, P.-J. Hsieh, and H.-Y. Wei, "Energy-efficient millimeter-wave M2M 5G systems with beam-aware DRX mechanism," in *Proc. IEEE 86th Veh. Technol. Conf. (VTC-Fall)*, Sep. 2017, pp. 1–5.

[12] C.-W. Weng, K.-H. Lin, B. P. S. Sahoo, and H.-Y. Wei, "Beam-aware dormant and scheduling mechanism for 5G millimeter wave cellular systems," *IEEE Trans. Veh. Technol.*, vol. 67, no. 11, pp. 10935–10949, Nov. 2018.

[13] C.-W. Weng, B. P. S. Sahoo, C.-C. Chou, and H.-Y. Wei, "Efficient beam sweeping paging in millimeter wave 5G networks," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC)*, May 2018, pp. 1–6.

[14] M. K. Maheshwari, M. Agiwal, N. Saxena, and A. Roy, "Hybrid directional discontinuous reception (HD-DRX) for 5G communication," *IEEE Commun. Lett.*, vol. 21, no. 6, pp. 1421–1424, Jun. 2017.

[15] M. Agiwal, M. K. Maheshwari, N. Saxena, and A. Roy, "Directional-DRX for 5G wireless communications," *Electron. Lett.*, vol. 52, no. 21, pp. 1816–1818, Oct. 2016.

[16] 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.

[17] M. K. Maheshwari, A. Roy, and N. Saxena, "Analytical modeling of DRX with flexible TTI for 5G communications," *Trans. Emerg. Telecommun. Technol.*, vol. 29, no. 2, p. e3275, Feb. 2018.

[18] N. R. Philip and B. Malarkodi, "Extended hybrid directional DRX with auxiliary active cycles for light traffic in 5G networks," *Trans. Emerg. Telecommun. Technol.*, vol. 30, no. 1, p. e3552, Jan. 2019.

[19] D. Liu, C. Wang, and L. K. Rasmussen, "Discontinuous reception for multiple-beam communication," *IEEE Access*, vol. 7, pp. 46931–46946, 2019.

[20] Z. Zhang, Q. Zhu, and P. Zhang, "Fast beam tracking discontinuous reception for D2D-based UAV mmWave communication," *IEEE Access*, vol. 7, pp. 110487–110498, 2019.

[21] E. Dahlman, S. Parkvall, and J. Skold, *5G NR: The Next Generation Wireless Access Technology*. New York, NY, USA: Academic, 2018.

[22] M. Lauridsen, G. Berardinelli, F. M. L. Tavares, F. Frederiksen, and P. Mogensen, "Sleep modes for enhanced battery life of 5G mobile terminals," in *Proc. IEEE 83rd Veh. Technol. Conf. (VTC Spring)*, May 2016, pp. 1–6.

[23] S. M. Ross, J. J. Kelly, R. J. Sullivan, W. J. Perry, D. Mercer, R. M. Davis, T. D. Washburn, E. V. Sager, J. B. Boyce, and V. L. Bristow, *Stochastic processes*, vol. 2., Wiley New York, 1996.

[24] *NR; Physical Channels and Modulation*, document TS 38.211, 3rd Generation Partnership Project (3GPP), 2019.

AN HUANG (Student Member, IEEE) received the B.S. degree in electrical engineering from National Taiwan University, Taipei, Taiwan, in 2018, where she is currently pursuing the M.S. degree in communication engineering. Since 2016, she has been working on DRX projects with the Wireless and Mobile Networking Laboratory led by Prof. H.Y. Wei. She was a Summer Intern with Foxconn, in 2017.

KUANG-HSUN LIN (Graduate Student Member, IEEE) received the B.S. degree in electrical engineering from National Taiwan University, Taipei, Taiwan, in 2015, where he is currently pursuing the Ph.D. degree in communication engineering with the Graduate Institute of Communication Engineering. Since 2015, he has been working with the Wireless Mobile Networking Laboratory led by Prof. H.Y. Wei. He held Summer Internships at Mediatek, in Summer 2015 and 2018.

HUNG-YU WEI (Senior Member, IEEE) received the B.S. degree in electrical engineering from National Taiwan University (NTU), Taipei, Taiwan, in 1999, and the M.S. and Ph.D. degrees in electrical engineering from Columbia University, New York, NY, USA, in 2001 and 2005, respectively.

He was a Summer Intern with Telcordia Applied Research, in 2000 and 2001. From 2003 to 2005, he was with NEC Laboratories America. He joined the Department of Electrical Engineering, National Taiwan University, in July 2005. He is currently a Professor with the Graduate Institute of Communication Engineering, Department of Electrical Engineering, National Taiwan University. He actively participates in wireless communications standardization activities. His research interests include broadband wireless communications, vehicular networking, cross-layer design for wireless multimedia communications, the Internet of Things, and game theoretic models for networking. From 2008 to 2009, he was a Consulting Member of the Acts and Regulation Committee of the National Communications Commission. He was a Voting Member of the IEEE 802.16 working group. He received the Recruiting Outstanding Young Scholar Award from the Foundation for the Advancement of Outstanding Scholarship, in 2006, the K. T. Li Young Researcher Award from the ACM Taipei Chapter and IICM, in 2012, the CIEE Excellent Young Engineer Award, in 2014, the NTU Excellent Teaching Award, in 2008, the Research Project for Excellent Young Scholars Award from Taiwan's Ministry of Science and Technology, in 2014, and the Wu Ta You Memorial Award from the Ministry of Science and Technology, in 2015. He was the Chair of the IEEE Vehicular Technology Society Taipei Chapter. He is the Chair of the IEEE P1935 standard working group on Edge/Fog Manageability and Orchestration.