Multi-RF Beamforming based Cellular Communication over Wideband mmWaves

Nancy Varshney and Swades De

Abstract

The existing literature on cellular multi-user mmWave communication focus on joint baseband and RF precoding designs to enable spatial multiple access with minimum interference. These studies either assume the number of users M is less than the number of RF units N_{RF} or schedule the users in time domain by dedicating one RF unit to each user if $M > N_{RF}$. It is expected that serving multiple users over OFDMA in each analog beam will offer better utilization of the wideband channel. To this end, for the scenario with $M \gg N_{RF}$ we propose a sectored-cell model that is supported by multi-RF chains over the wideband mmWave channel, with each beam serving multiple users within a sector and the sectors being scheduled in round-robin fashion. We also propose a variable time frame structure that conducts sector-wise initial access, with simultaneous access to all users within a sector. It provides improved spectrum efficiency and decreased initial access delay as compared to the initial access using exhaustive beam search method. We then jointly estimate the optimum beamwidth and optimum N_{RF} that offer maximum average long-run user rate. Further, we introduce a reduced-complexity sector sojourn time optimization for non-homogeneously distributed users, that improves fairness of long-run user rates leveraging the variable time frame structure. The numerical results show that, while a high value of N_{RF} causes more interference to peak data rate, the average long-run user rate improves. Additionally, using a very narrow beam is also not optimal for providing maximum rate support. The proposed beamforming method offers a higher average long-run user rate over the competitive beamforming schemes while the complexity of user scheduling is independent of M.

Index Terms

interference, millimeter wave (mmWave) communications, multi-RF, optimal beamwidth, sector sojourn time, scheduling, user distribution

A preliminary version of the work was presented in 27th National Conference on Communications, 2021 [1].

I. INTRODUCTION

Due to high spectrum availability at mmWaves there has been a lot of interest in directional mmWave wireless communication studies at the physical and medium access control (MAC) layer recently. Deployment of massive massive multiple-input multiple-output (MIMO) is envisioned at mmWaves. It is imperative to reduce the hardware and system complexity of a large mmWave system. As a result, various hybrid precoding designs to maximize sum rate while achieving a reduction in the number of RF units, inter-user interference have been investigated. However, this digital-analog beamforming requires two-step non-polynomial time (NP)-hard precoding, namely baseband and RF precoding leading to high computational complexities.

The mmWave communication systems can be deployed in two modes, namely, non-standalone (NSA) mode – where control signaling is provided by existing 4G networks, and standalone (SA) mode – where control signaling is also carried out at mmWave frequencies. Though much research has been reported on the NSA domain, the SA mode has received relatively much less attention. SA mode will help mmWaves to realize its full potential in terms of ultra-low latency and huge bandwidth. On the other hand, the SA mode has its own set of difficulties. One such challenge is to identify a suitable narrow beam pair between eNodeB (eNB) and user equipment (UE) during the initial access (IA) phase. The Third Generation Partnership Project (3GPP) Fifth Generation New Radio (5G NR) [2] has defined numerous standards for SA cellular mmWave communications, which are still being considered. As a result, the advancement of mmWave beamforming technologies and numerous outdoor channel measurement campaigns [3], [4] encourages to probe into mmWave outdoor commercial cellular systems, especially where user density per cell is sufficiently high.

A. Related work and motivation

Several works have focused on formulating hybrid precoding algorithms for a single user, multi-user, and MIMO mmWave systems, over narrowband channels as well as wideband frequency-selective channels. The authors in [5] suggested allocating the same UE over the wideband channel by assuming a high correlation among subchannels due to the sparse nature of the mmWave channel. However, due to the beam-squinting effect of large antenna arrays and the surrounding environment, this is not always true [6]. Further, mmWave channels have a frequency-selective nature because of the huge bandwidth range (giga-Hertz range). Till now, limited work has been done on the high UE population wideband mmWave hybrid beamforming scenario.

The literature on multi-user hybrid beamforming have mainly focused on using spatial division multiple access (SDMA) by obtaining orthogonality on data streams to UEs via joint RF and baseband precoding schemes based on channel information and correlation. For wideband mmWave channel, the authors in [7] designed a hybrid precoder using single-carrier and assigned different numbers of data streams to the UEs equipped with multiple RF chains. The authors in [8] proposed a subchannel constrained fully-connected hybrid precoding design over OFDM symbol with a limit on the total number of UEs, thereby requiring $N_{RF} \geq M$. However, the computation complexity of this design is very high. Moreover, the optimal UE assignment over a subchannel is assumed to be known a priori. Further, when $M \gg N_{RF}$, to maximize the multi-user diversity sub-optimal SDMA schemes are considered with user scheduling in time domain. In the case of $M > N_{RF}$, the approaches in [9], [10] grouped the UEs at RF beamforming stage based on the channel's second-order statistics and the angle-of-departure (AoD) similarities, respectively, and then using baseband precoder multiplexed them over dominant channel eigenmodes. In [11] authors proposed grouping and scheduling of UEs with reduced complexity employing sub-array structure. In [12] the authors focused on codebook design with limited feedback to schedule N_{RF} UEs at a time and showed that RF-only beamforming can outperform hybrid beamforming with reduced inter-user interference (IUI). The limitations of schemes in [9]–[12] is that these schedule a subset of UEs with good channel conditions only for each transmission in time domain. Hence, dedicating a data stream generated from an RF unit to a single UE at a time over a wideband mmWave channel will lead to high scheduling complexities and poor system performances.

It is notable that, if two UEs are co-located then even with baseband precoding, multiplexing UEs over dominant eigenmodes, perfect orthogonality might not always be possible, resulting in IUI. However, if the UEs are multiplexed over orthogonal subcarriers, IUI does not arise. Further, to boost the system capacity when $M \gg N_{RF}$, a wideband mmWave channel can be utilized to schedule multiple UEs over a single beam using OFDMA. To serve multiple UEs in OFDMA, the authors in [13] used sparsity-based fully digital precoding to jointly estimate subcarrier and hybrid precoder design for a fully-connected hybrid architecture. However, even with weighted minimum mean squared error digital (MMSE) precoder in [13], perfect orthogonality among the spatial streams is not always possible in the case of co-located UEs. Besides, with increasing UE population and antenna array size, the complexity of zero-forcing and MMSE baseband precoders grow non-linearly. Moreover, zero-forcing and MMSE precoders are both constrained by the weakest UE signal-to-noise ratio (SNR). To this end, in our preliminary studies [1] we developed

an SDMA scheme for an eNB deployed with multi-RF chains to schedule $M \geq N_{RF}$ UEs at a time using analog beamforming dependent sectored cell model. Subsequently, at baseband level inter-beam interference is minimized at OFDM resource allocation stage.

According to 3GPP NR guidelines, the communication at mmWaves is divided into two phases - beam training phase and data transmission phase. During beam training, the eNB performs an IA to determine the optimal eNB-UE beam pair and to acquire additional UE parameters. The beam training occurs periodically to update channel and UE related information. Several approaches for IA, namely, exhaustive search, hierarchical search, and context-information-based search, have been investigated in the literature for mmWave systems [14]–[16]. The authors in [17] compares the latency of different combinations of directional beamforming and omnidirectional transmission IA protocols. A multi RF-chain sub-array technique is proposed in [18] for optimizing hierarchical codebook using beam widening. Since hierarchical search is better suited for point-to-point indoor communications, brute exhaustive search is the IA method suggested for 5G NR cellular scenarios [19]. The authors in [19] presented a survey on 3GPP NR MAC layer design at mmWave frequencies and studied the implications of its various parameters on IA and beam training. However, the two-phase communication results in under-utilization of huge mmWave bandwidth because the beam training data is usually very small. Further, in 3GPP NR a slot is proposed to be composed of 14 OFDM symbols for enhanced mobile broadband (eMBB) applications at the mmWave range [20]. This fixed frame structure again results in resource under-utilization, specially in beam training phase, and fairness discrepancies among UEs in the mmWave systems. In [21], the authors proposed a directional mmWave communications using flexible frame structure based on variable transmission time interval (TTI) and proved that variable TTI offers various advantages, such as reduced latency and flexible scheduling. Table I summarizes the existing mmWave MAC layer protocols.

To this end, we observe that in a multi-user scenario assigning the entire wideband mmWave channel to the same UE is not optimal. In our prior work in [25] we studied the sectored cell mmWave system with a single RF unit using RF-only precoding, to serve multiple UEs at a time over OFDMA, thus bypassing baseband zero-forcing precoding. However, since it was based on the assumption of a single beam, no inter-beam interference analysis was performed, and the MAC layer analysis was not considered. In [1] we studied the effect of the number of beams and number of sectors on the average long-run UE rates of homogeneously distributed UEs. As an advance, in this paper, we investigate the inter-beam interference for different UE distributions

Table I: Existing MAC protocols for standalone mmWave systems

References	Frame flexibility	UE Scheduling	IA	Limitations
IEEE 802.11ad [22], [23]	fixed slot boundaries	single UE per RF beam; multiple UEs scheduled in spatial and time domain	hierarchical search	IA delay increase with narrow beams at eNB and UE; not suitable for cellular communications
3GPP NR [19], [24]	flexible uplink (UL) and DL resource allocation in a slot; fixed slot boundaries	not specified	exhaustive search	significant IA delay; bandwidth under-utilization during beam training phase
Frame structure [21]	flexible UL and DL resource allocation in a slot; fixed slot boundaries but flexible TTI	single UE per RF beam at a time	exhaustive search	variable TTI optimized only for short packets transmission; dedicated TDMA scheduling results in bandwidth under-utilization; assumes perfect beam alignment

using a sectored-cell approach with multiple RF units. Further, the existing MAC layer protocols at mmWaves are not spectrally efficient in the IA phase. To this end, we propose the related MAC layer specifications that provide simultaneous IA to all the UEs within a sector.

B. Contribution

In this paper we study the performance of the sectored-cell system model with multi-RF beams while accounting for the resulting inter-beam interference and UE distribution dependent variable sector sojourn time, to cater to a large UE population. The performance metrics of the proposed system depend on the optimal number of sectors as well as on selection of the optimal number of concurrently active beams. To the best of our knowledge, the proposed system model has not been studied in the literature yet. The key contributions are:

1. A novel sectored-cell system model is presented using a multi-RF beamforming structure to serve the UEs in the urban micro (UMi) scenario at mmWave frequency. Each RF unit is connected to a array of an optimal size capable of forming narrow steerable beams over the sectors. Each beam supports multiple UEs per sector at a time. At the baseband level, algorithms are proposed to optimally allocate OFDM resources to the UEs to minimize inter-beam interference while maximizing throughput. Further, the two schemes namely, synchronous and asynchronous sweeping of cell by the beams are investigated.

2. A variable time frame structure is proposed for the sectored-cell model. It aims to reduce IA delay and achieves better spectral efficiency compared to exhaustive search IA procedure proposed by 3GPP NR for cellular mmWave communications.
3. For a given eNB power budget and a large number of homogeneously distributed UEs in the cell, a joint optimization problem is formulated to find the optimum sector beamwidth and an optimum number of concurrent beams that maximize the average long-run UE rate in presence of side-lobe interference using realistic uniform linear antenna array (ULA) gain pattern. Also, the comparison of user scheduling complexity of the proposed system is provided over the existing competitive scheduling models in the literature.
4. In the case of non-homogeneous UE distribution, sector sojourn time is optimized by availing the advantages of variable time frame structure using weighted round-robin (WRR) scheduling to improve the fairness of long-run UE rates.

C. Organization

The system model is presented in Section II. The variable time frame structure is presented in Section III. Resource allocation strategy, optimum sector beamwidth, number of RF units, and sector sojourn time are analyzed in Section IV. Complexity analysis is presented in Section V, simulation results are discussed in Section VI, and concluding remarks are drawn in Section VII. **Notations:** $|\cdot|$ and $\mathcal{O}$ denote cardinality and algorithm run-time complexity, respectively.

II. System Model

In this section, we describe the cell architecture and the antenna configuration at the eNB to support the directional communication at the mmWave in the UMi environment, with the number of RF chains $N_{RF} \ll M$, where M is the number of UEs independently distributed in the cell. The cell is partitioned into S identical sectors. We also propose an underlying variable time MAC layer protocol to support the proposed sectored-cell time-multiplexed model in section III.

Let $\mathcal{S}$ be the set containing all possible values of S that a cell can be sectorized into, i.e., $\mathcal{S} = \{S | S = \lceil \frac{2\pi}{\Theta} \rceil, \Theta_{min} \leq \Theta \leq \Theta_{max}\}$, where $[\Theta_{min}$ and $\Theta_{max}]$ denotes the range of supportable sector width or equivalently half power beamwidth (HPBW) of the serving beam. For a particular sectorization scheme S, let s denote the index of a sector, $s = \{1, 2, \cdots, S\}$. Also, let K_s be the number of UEs located in sth sector of a sectorization scheme S such that $\sum_{s \in S} K_s = M$.

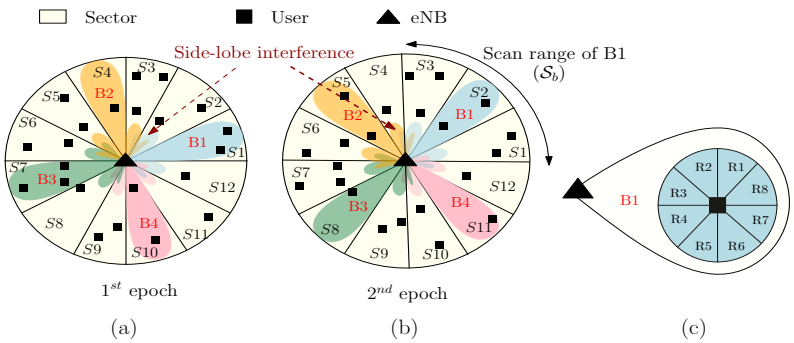


Figure 1: (a) and (b) depicts round-robin scheduling of sectors over two consecutive epochs with $S = 12$ sectors and $N_{RF} = 4$ beams, represented as $\{S1, \cdots, S12\}$ and $\{B1, \cdots, B4\}$, respectively. Here, $\mathcal{S}_{b=B1} = \{S1, S2, S3\}$; and (c) shows steering directions $\{R1, \cdots, R8\}$ of a UE within beam B1, i.e., $S_r = 8$.

A. UE deployment

In this paper we study the performance in context to both homogeneous and non-homogeneous UE deployment. In homogeneous UE setup, the probability of a UE falling in a sector s is given as $\mathcal{Q}(s) = 1/S$. To simulate non-homogeneous binomial point process divide a cell into Z zones. Further, each zone has a granularity of $\delta\phi$. The probability of a UE falling in the angular area $ar(\hat{\phi})$ varies with $\hat{\phi}$, where $\hat{\phi}$ is the angular distance. Then, $\mathcal{Q}(ar(\hat{\phi}))$ is calculated as

$$\mathcal{Q}(ar(\hat{\phi})) = \frac{\exp(\Upsilon \lfloor \text{mod}(\hat{\phi}, 360^o/Z)/\delta\phi \rfloor \delta\phi)}{Z \sum_{i=0}^{(360^o/Z\delta\phi)-1} \exp(\Upsilon i \delta\phi)} \tag{1}$$

where Υ decides the intensity of non-homogeneity. The proof of $\int_{\hat{\phi}} \mathcal{Q}(ar(\hat{\phi}))d\hat{\phi} = 1$ is given in Appendix A. The distribution given by (1) is used in Section VI-C for simulation studies.

B. Spatial multiplexing with N_{RF} beams

Spatial multiplexing is accomplished at the eNB by scanning the 2π angular space with N_{RF} RF chains that concurrently serve UEs located in N_{RF} out of S sectors in an epoch (Fig. 1 (a) and (b)). We assume that S is a multiple of N_{RF}. Each RF unit is connected to a separate array of N_t antenna elements. Each beam serves all the UEs in a sector using OFDMA at a time. Thus, the N_{RF} beams serve the sectors in a fashion such that each beam steers over a set of S/N_{RF} consecutive sectors in a cell sweep. Let s_j denote sth sector served by jth beam in an epoch. Then, $\mathcal{S}_b = \{s_b | s \text{ served by the } b\text{th beam}\}$, where $b = 1, \cdots, N_{RF}$. Further, we assume that each UE has a single RF unit connected to array of N_r antenna elements. For analytical simplicity, we consider fixed beamforming gain $G_r = N_r$ at the receiver irrespective of the steering directions, and let number of steerable directions at UE be S_r, as shown in Fig. 1 (c).

C. Wideband mmWave channel model

At mmWave frequencies there exists either a line-of-sight (LOS) link or non-line-of-sight (NLOS) link between UE and eNB due to large difference in attenuation of first order and higher order signals. The probability of a link being in LOS for kth UE at distance d_k from eNB is expressed as $\Pr(d_k) = \min(d_1/d_k, 1)(1 - e^{-d_k/d_2}) + e^{-d_k/d_2}$, where $d_1 = 18$ and $d_2 = 36$ [26]. Additionally, the path loss for LOS and NLOS link are obtained as [27]

$$\begin{aligned} \text{PL}_{LOS}(d_k) &= 61.4 + 20\log_{10}(d_k) + \mathcal{N}(0, 33.64) \text{ [dB]} \\ \text{PL}_{NLOS}(d_k) &= 72.0 + 29.2\log_{10}(d_k) + \mathcal{N}(0, 75.69) \text{ [dB]}. \end{aligned} \tag{2}$$

Therefore, for the kth UE, at a distance d_k from eNB the pathloss PL_k is expressed as

$$\text{PL}_k = \begin{cases} \Pr(d_k) 10^{\text{PL}_{LOS}(d_k)/10}, & \text{LOS} \\ (1 - \Pr(d_k)) 10^{\text{PL}_{NLOS}(d_k)/10}, & \text{NLOS}. \end{cases} \tag{3}$$

The channel between the eNB and the kth UE exhibit frequency-selective fading; the total available bandwidth B is divided into N_c subchannels, with f_c being the carrier frequency. A subchannel consists of multiple subcarriers. Further, the channel consists of $L_k \ll N_t$ multipath components (MPCs) [28]. Thus, the channel between the eNB and the kth UE over the nth subchannel (frequency f_n) as seen by a single RF unit connected to an ULA of N_t active antenna elements is expressed as

$$\mathbf{h}_{k,n} = \sqrt{\frac{N_t}{L_k \text{PL}_k}} \sum_{l=1}^{L_k} \alpha_{k,n,l} \mathbf{a}_{\mathbf{T}}(\Omega(f_n), \phi_{k,l})^H \in \mathbb{C}^{1 \times N_t} \tag{4}$$

where $\alpha_{k,n,l}$ is small scale fading gain of lth MPC over nth subchannel, and $\phi_{k,l}$ is the AoD of lth MPC, assumed to uniformly distributed with a angular spread of σ_T^{AS}. The eNB array response vector $\mathbf{a}_{\mathbf{T}}(\Omega(f_n), \phi_{k,l})$ at an offset angle $\phi_{k,l}$ is given as

$$\mathbf{a}_{\mathbf{T}}(\Omega(f), \phi_{k,l}) = \frac{1}{\sqrt{N_t}} \left[1, e^{-j\frac{2\pi}{\lambda_c} d' \Omega(f) \sin\phi_{k,l}}, e^{-j\frac{2\pi}{\lambda_c} d' \Omega(f)(N_t-1)\sin\phi_{k,l}}\right]^T \tag{5}$$

Here, $\Omega(f) = (1 + f/f_c)$ is the beam squint parameter at frequency f [6], d' is the inter element ULA spacing, and λ_c is the carrier wavelength. The information of $(L_k, \alpha_{k,l}, \phi_{k,l}, d_k)$ $\forall k$ is available at the eNB from the IA phase assuming zero estimation error.

Let G_0 be the gain of a single antenna element. Then, the maximum sub-array gain with N_t antenna elements is $G_\Theta = N_t G_0$ with HPBW $\Theta \approx 2/N_t$ [29]. In this system model, by appropriate power control, it is ensured that the beam coverage range is constant irrespective of beamwidth. For this, we fix the effective isotropic radiated power (EIRP) to a value that

is sufficient to compensate for both high attenuation and high noise power due to large signal bandwidth at mmWave frequencies and is also below the maximum limit specified in Federal Communications Commission (FCC) guidelines [30]. Now, consider P_{total} as the total transmit power per RF chain required corresponding to beamwidth Θ_{max} for a specified EIRP level. Then, for a beamwidth $\Theta < \Theta_{max}$ the transmit power required to guarantee the same radiated power density at distance d is calculated as

$$P_t = \frac{\text{EIRP}}{G_\Theta} = \frac{P_{total} G_{\Theta_{max}}}{G_\Theta} \tag{6}$$

where $G_{\Theta_{max}}$ and G_Θ are the broadside ULA gain corresponding to Θ_{max} and Θ, respectively.

D. Effective rate in presence of side-lobe interference

In an epoch out of total S sectors only N_{RF} sectors will be served simultaneously. So, with the consideration of practical beam radiation pattern with side-lobes at eNB, the UEs in s_bth sector served by bth beam in an epoch, where $b = \{1, 2, \cdots, N_{RF}\}$, will experience side-lobe interference from the rest of the beams $\hat{b} \neq b$. Also, at eNB let $\mathbf{A}$ be the wideband analog beamforming vector of a beam b, steered at offset angle Φ_b to serve sth sector, given as

$$\mathbf{A}(\Phi_b) = \frac{1}{\sqrt{N_t}} \left[1, e^{-j\frac{2\pi}{\lambda_c} d' \sin\Phi_b} \cdots, e^{-j\frac{2\pi}{\lambda_c} d'(N_t-1)\sin\Phi_b}\right]^T. \tag{7}$$

In a sectorized cell with total S number of sectors, the possible steering directions are fixed at $\Phi_b = (t + b - 2)(360^o/S)$, where $t = 1, \cdots, \lceil S/N_{RF} \rceil$ denotes the epoch index during a cell sweep. Consequently, the signal-to-interference-plus-noise ratio $\Gamma_{k,n}^{s_b}$ of kth UE at distance d_k from eNB, over the nth subchannel in sector s_b served by beam b is

$$\Gamma_{k,n}^{s_b} = \frac{P_{k,n}^{s_b} |\mathbf{h}_{k,n}^{s_b} \mathbf{A}(\Phi_b)|^2 G_r}{N_0 B/N_c + \sum_{j=1, j\neq b}^{N_{RF}} P_{k,n}^{s_j} G_r |\mathbf{h}_{k,n}^{s_j}{}^H \mathbf{A}(\Phi_j)^H \mathbf{A}(\Phi_j) \mathbf{h}_{k,n}^{s_j}|} = P_{k,n}^{s_b} \gamma_{k,n}^{s_b} \tag{8}$$

where N_0 is the noise spectral density, and $P_{k,n}^{s_j}$ and $\mathbf{h}_{k,n}^{s_j}$ are the power allocation and channel vector of kth UE over nth subchannel, respectively, in s_jth sector served by jth beam. Thus, data rate of kth UE over the nth subchannel in s_bth sector in an epoch is $r_{k,n}^{s_b} = (B/N_c)\log_2(1+\Gamma_{k,n}^{s_b})$.

E. RF chain power consumption

At mmWaves, the power dissipation in RF circuitry is also a significant part of the total eNB power budget. The power consumption values and quantities of all the components in an RF front-end are listed in Table II [31]. Since the power consumption of passive phase-shifter is negligible, the total power consumption per RF chain, having a separate I/Q module,

Table II: Power consumption of components in a RF chain

Component	Notation	Power consumption	Component	Notation	Power consumption
DAC	P_{DAC}	Given by (9)	Low-pass filter	P_{LPF}	14 mW
Mixer	P_M	0.3 mW	Phase-shifter	P_{PS}	≈ 0 mW
Local oscillator	P_{LO}	22.5 mW	Power amplifiers	P_{PA}	$P_t/\eta_{PA}, \eta_{PA} = 27\%$

is $P_{rf}(P_t) = 2P_{DAC} + 2P_M + 2P_{LPF} + P_{LO} + P_{PA}$. Here, P_{PA} depends on transmit power P_t and power amplifier's efficiency η_{PA}. P_{DAC} depends on the sampling frequency F_s and number of bits b_t as

$$P_{DAC} = 1.5 \times 10^{-5} \cdot 2^{b_t} + 9 \times 10^{-12} \cdot b_t \cdot F_s. \tag{9}$$

An eNB employing N_{RF} number of RF chains will have a single local oscillator shared among all the chains. Therefore, total power consumption of RF front-end at eNB is

$$P_{nrf}(N_{RF}, P_t) = N_{RF}(2P_{DAC} + 2P_M + 2P_{LPF} + P_t/\eta_{PA}) + P_{LO}. \tag{10}$$

Remark 1. *The primary power-hungry component in an RF chain is the power amplifier. The energy consumption of a power amplifier depends on the transmitted power P_t as well (as listed in Table II). Therefore, for a given value of N_{RF}, decreasing beamwidth will decrease the transmit power requirement (from* (6)*), and hence it will reduce the resulting power waste in power amplifiers.*

III. Proposed Variable Time Frame Structure

In this section, we propose a variable time frame structure that offers greater flexibility for control signaling and sector scheduling by a beam. The proposed frame structure has smaller IA delay and control overhead compared to the IA procedure proposed in 3GPP NR [32].

Consider two states of UE operation: idle state and connected state. In an idle state, a UE does not communicate with eNB and only passively monitor eNB's paging and broadcast information to maintain connectivity with the network. The UE transits from idle to connected state after the initial handshake procedure. The proposed sectored-cell system model uses N_{RF} concurrent beams for spatial multiplexing and each spatially distributed beam serves only a subset of $\mathcal{S}_b$ sectors, as shown in Fig. 1(a). Therefore, in the following section, we consider the case of a single analog beam covering $\mathcal{S}_b$ sectors in a round-robin fashion in a single cell sweep. Furthermore, each UE has a single RF chain connected to a ULA capable of forming narrow beam steerable

in S_r directions, as stated in section model Section II. The same procedure occurs independently with the rest of the concurrently active analog beams, achieving full spatial multiplexing.

A. 3GPP NR IA procedure

The smallest 3GPP 5G NR frame is equal to one slot duration which is also equal to its TTI. A TTI is a set of OFDM symbols that constitutes the smallest decodable transmission unit. Each slot can carry a maximum of 2 synchronization signal (SS) blocks that are used for beam training and control information broadcast. Each SS block comprises 4 consecutive OFDM symbols over a set of 240 subcarriers [33]. The IA steps of a new UE during the beam management as in 3GPP NR Release 15 are shown in Fig. 2(a).

The IA procedure of 3GPP NR is as follows. First, the beam sequentially scans all of the $\mathcal{S}_b$ sectors and transmits SS blocks corresponding to each eNB-UE beam pair directions, i.e., totaling $|\mathcal{S}_b|S_r$ directions. Based on maximum SNR, the new UE selects the best beam and decodes the corresponding SS block while the connected UEs perform beam tracking and paging in the rest of the control portion. The UE obtains essential system information by decoding primary synchronization signals, secondary synchronization signals, and physical broadcast control channel after acquiring SS blocks. The UE then connects to the network using a physical random access channel (PRACH) to send random access (RACH) preambles in the UL slot [32]. RACH preamble from a UE informs the eNB about its channel and beam conditions. In standalone mmWave, an extra round of sequential scan over $\mathcal{S}_b$ is performed to notify UEs about the unique time-frequency offset of PRACH corresponding to each SS block transmitted in all eNB-UE pair directions for best beam estimation. It allows UEs to know when and where to transmit their RACH preamble for contention-free access, thus extending the IA.

B. Proposed IA procedure

We perform beam training and data transmission simultaneously within a frame using the proposed frame structure, leveraging the sectored-cell model. Some resources in the frame are set aside for IA of new UEs, while the remaining ones are used for data transmission with connected UEs. Therefore, jointly with variable time structure of the frame and combined beam training and data transmission phase, the mmWave bandwidth is efficiently utilized.

In the proposed variable time frame structure we assume a superframe of duration T_{tot}. The superframe is composed of frames transmitted by the beam in all the sectors $s \in \mathcal{S}_b$ in a

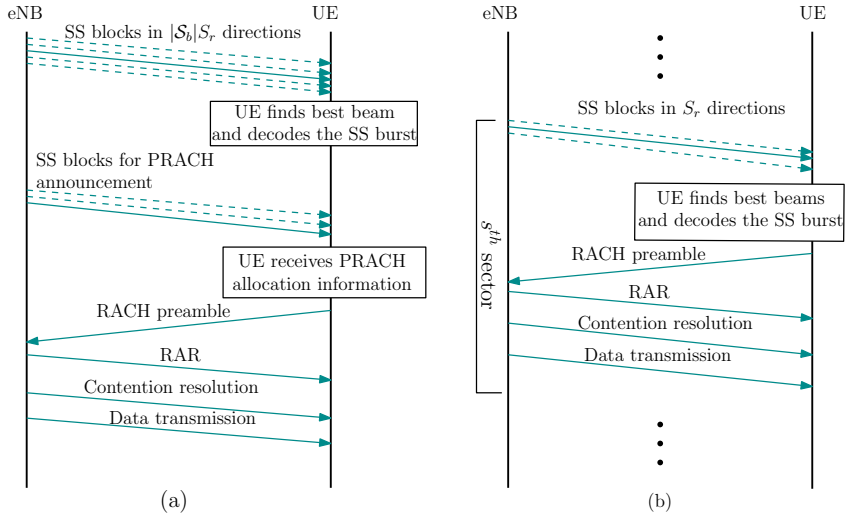


Figure 2: (a) Steps in IA for new UEs using 3GPP NR guidelines [19] (b) Steps in IA for news UEs according to proposed frame structure for directional cellular communication.

single cell sweep. There are no fixed frame boundaries, i.e., each frame in a sector can have a different number of OFDM symbols instead of a fixed duration of 14 OFDM symbols in 3GPP NR. Further, as explained above, the beam management of new UEs and exchange of data with multiple connected UEs occur simultaneously using OFDMA in the control portion of the frame. Subsequently, we consider that the frame in a sector $s \in \mathcal{S}_b$ consists of a control subframe (beacon) of duration τ plus a subframe of variable duration T_s, as shown in Fig. 3, such that $\sum_{s\in\mathcal{S}_b}(\tau + T_s) = T_{tot}$. The beacon duration τ is the same for all sectors.

The paging queries for the connected UEs are included in the SS burst while the paging responses by the UEs are transmitted over separate UL channels. Fig. 2(b) depicts the steps required in IA procedure. For brevity, we describe only the modified channels, while keeping the functionality of other channels same. Furthermore, different TTI duration can be defined using the different number of OFDM symbols. In the proposed frame structure we set the minimum TTI possible equal to one OFDM symbol. Hence, the UEs can capture and decode the information of one TTI without waiting for the entire frame to end.

The IA using the proposed variable time frame structure is explained as follows. As the beam b enters a sector s_b, it broadcasts SS blocks in the first $4S_r$ OFDM symbols for the duration t_{c1} of the control subframe, allowing new UEs to search for the SS block in each of its S_r directions and decode the one with the strongest signal. Moreover, with the proposed system the likelihood of RACH collision from multiple new UEs in a sector is negligible.

Lemma 1: When N_{RF} is sufficiently high, the probability of two or more UEs transmitting

the same RACH preamble over PRACH in a single frame is very small.

Proof: See Appendix B.

Consequently, multiple time-frequency resources corresponding to different S_r directions are not set aside in PRACH, thereby saving bandwidth. Then the new UEs transmit orthogonal RACH preambles over a single set of reserved time-frequency resources assigned for PRACH in the next TTI of 1 OFDM symbol duration (t_{c3} duration) in UL after waiting for t_{c2} duration. The RACH also contains the information of the UEs' best link and channel information. If there is contention because of two or more UEs using the same preamble over the same time-frequency grid, eNB is still able to detect the presence of contending UEs. The eNB decodes RACH preambles to access UEs beam conditions, namely, gain, steering direction, etc., and accordingly finds the optimum resource allocation strategy. After waiting for t_{c4} duration, eNB grants UEs the random access response (RAR) over the next TTI consisting of 2 OFDM symbols (t_{c5} duration) in the common physical DL control channel (PDCCH). The transmission of RAR to a UE indicates that the eNB has correctly obtained and decoded the RACH preamble, allowing the UE to access the channel. The UE can now access the channel immediately after τ duration. As a result, the complete IA of all the UEs located in a sector can be carried out within the sector sojourn time. Here, t_{c2} and t_{c4} are the time offset to allow for transmission and processing delay. The remaining resources in τ duration are used for data transmission to connected UEs

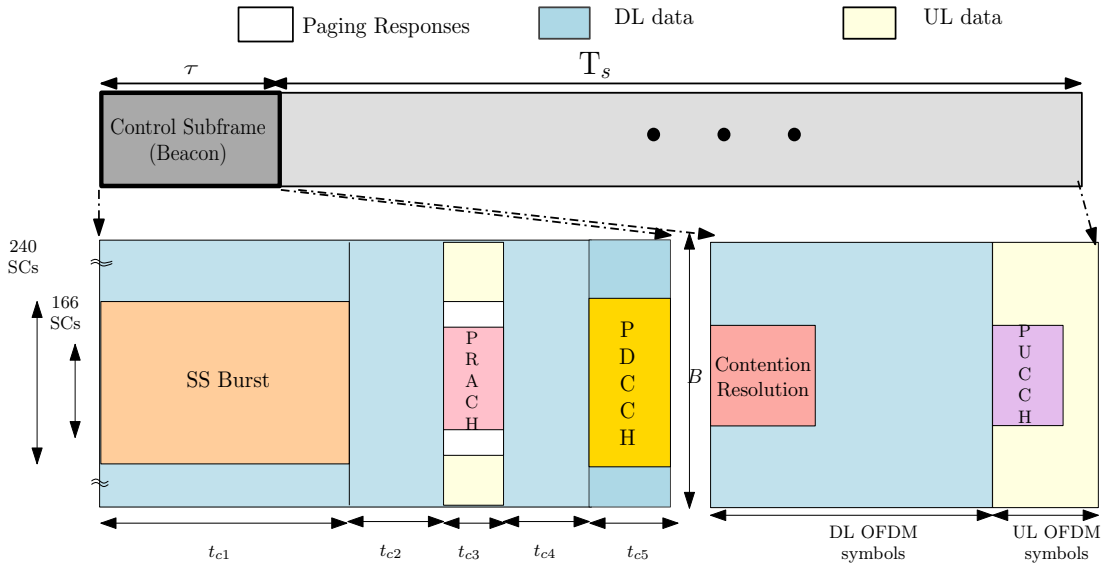


Figure 3: Schematic diagram of the proposed frame structure, transmitted in sth sector, depicting the position of control channels required for IA in τ duration, and control channels required for contention resolution and UE beam refinement in T duration.

in DL/UL. If in sth sector $K_s = 0$ then eNB does not detect any paging responses and RACH preambles, the frame is of duration τ only. The duration T_s is dedicated for data portion, which can be optimized to the granularity of OFDM symbol to improve UEs' fairness, as discussed later in Section IV-C. In addition, as shown in Fig. 3, DL and UL assignment in a frame is flexible, and some subcarriers are reserved for contention resolution in PDCCH and physical UL control channel DL (PUCCH) in T_s duration.

C. Performance metrics

1) IA delay: In directional communication IA delay (IAD) is defined as the time required by a new UE to transit into the connected state after it first listens to the synchronization signals. According to the methodology defined by 3GPP NR, the IA procedure takes place by scanning all the $\mathcal{S}_b$ sectors sequentially for SS burst and then reception of RACH in $\mathcal{S}_b$ sequentially before sending RAR to all the new UEs. Also, 3GPP NR defines TTI size equal to one slot and supports a maximum of two SS blocks per slot [32]. Thus, the total time required is summation of $(|\mathcal{S}_b|S_r)/2 \times 14T_{OFDM}$ for SS block transmission, $|\mathcal{S}_b| \times 14T_{OFDM}$ for PRACH announcement, $|\mathcal{S}_b| \times 14T_{OFDM}$ for RACH transmission, and $|\mathcal{S}_b| \times 14T_{OFDM}$ for RAR. Therefore, minimum IAD of procedure defined in 3GPP NR is given as, $\text{IAD}^{3GPP} = |\mathcal{S}_b| \cdot 14 \cdot (S_r/2 + 3)T_{OFDM}$.

On contrary, in the proposed variable time frame structure allows for the IA to be carried out sector by sector. Therefore, compared to the 3GPP NR, minimum IAD of the proposed scheme is equal to the beacon subframe time, i.e., $\text{IAD}^{Proposed} = (4 \cdot S_r + 3)T_{OFDM} + t_{c2} + t_{c4} = \tau$. Thus, the percentage improvement in IAD with the proposed variable time frame structure is

$$\Delta \text{ IAD} = \left(1 - \frac{\tau}{|\mathcal{S}_b| \cdot 14 \cdot (S_r/2 + 3)T_{OFDM}}\right) \times 100. \tag{11}$$

2) Bandwidth saving: From Lemma 1 we know that, unlike the IA methodology in 3GPP NR, in the proposed IA procedure there is no need to reserve different resource elements for different UEs over the bandwidth for RACH reception, thereby saving bandwidth. Consequently, the proposed IA method does not need additional rounds of SS block transmission for RACH resource announcement in $\mathcal{S}_b$ directions. The SS block uses 240 subcarriers over 4 OFDM symbols, RACH preamble occupies a maximum of 166 subcarriers, and RAR transmitted on PDCCH that consumes a maximum of 240 subcarriers [20]. Hence, the total bandwidth consumed in overhead for the IA procedure using 3GPP NR specifications (assuming no contention) is

$$\text{BW}^{3GPP} = \Delta f \cdot \text{REs in}\{\text{SS blocks} + \text{RACH resources announcement} + \text{RACH preambles} + \text{RAR}\}$$

$$= \Delta f \cdot (240 \cdot 4 \cdot |\mathcal{S}_b| \cdot S_r \cdot N_{rep} + 240 \cdot 4 \cdot |\mathcal{S}_b| + 166 \cdot |\mathcal{S}_b| + 240 \cdot 2 \cdot |\mathcal{S}_b|) \tag{12}$$

where RE is the resource element (smallest time-frequency element), Δf is the subcarrier spacing and N_{rep} is total number of SS blocks transmitted over bandwidth B per eNB-UE beam pair. The total bandwidth required in IA procedure using proposed frame structure is

$$\begin{aligned} \text{BW}^{Proposed} &= \Delta f \cdot \text{REs in } \{\text{SS blocks + RACH preambles + RAR}\} \\ &= \Delta f \cdot (240 \cdot 4 \cdot |\mathcal{S}_b| \cdot S_r + 166 \cdot |\mathcal{S}_b| + 240 \cdot 2 \cdot |\mathcal{S}_b|). \end{aligned} \tag{13}$$

Hence, the percentage saving in the bandwidth using the proposed frame is

$$\Delta \text{ BW} = \frac{480 \cdot (S_r \cdot N_{rep} - S_r + 1)}{480 \cdot S_r \cdot N_{rep} + 803} \times 100. \tag{14}$$

IV. Problem Formulation and Performance Evaluation

In this section, we evaluate the performance of the proposed system with the underlying protocol as presented in Section III in terms of average long-run UE rate, power budget, and fairness for homogeneously distributed UEs. The analysis is further extended to find the optimal sector sojourn time in case of non-homogeneous UE distribution for a given S^* and N^*_{RF} that optimizes system performance.

For the ease of analysis, we assume the same DL and UL data capacities and calculate the total data rate supported over the bandwidth. We also assume quasi-static channel during a cell sweep. Suppose N_{c1} and N_{c2} are the numbers of effective subchannels available for data transmission during τ and T_s interval, respectively. We define the long-run UE rate as the total data bits received by a UE over one complete cell sweep duration T_{tot}. Therefore, the long-run rate of kth UE located in sth sector is

$$R^s_k = \frac{\tau_s \sum_{n=1}^{N_{c1}} \pi^{s,\tau_s}_{k,n} r^{s,\tau_s}_{k,n} + T_s \sum_{n=1}^{N_{c2}} \pi^{s,T_s}_{k,n} r^{s,T_s}_{k,n}}{T_{tot}} \tag{15}$$

where τ_s and T_s are the control subframe and variable subframe duration of sth sector and $\tau_s = \tau$ $\forall s$. The variables $\pi^{s,\tau}_{k,n}$ and $\pi^{s,T_s}_{k,n}$ denote the subchannel assignment in OFDM symbols during τ and T_s, respectively. The value of $\pi^{s,\tau/T_s}_{k,n} = 1$ denotes the assignment of nth subchannel to kth UE is sth sector during τ or T_s. Similarly, we define the average long-run UE rate $\bar{R}$ as the average of long-run rate of all UEs and it is equivalent to

$$\bar{R} = \frac{\sum_{s=1}^{S} \sum_{k=1}^{K_s} R^s_k}{M} = \frac{1}{MT_{tot}} \sum_{s=1}^{S} \left(\tau \sum_{k=1}^{K_s} \sum_{n=1}^{N_{c1}} \pi^{s,\tau}_{k,n} r^{s,\tau}_{k,n} + T_s \sum_{k=1}^{K_s} \sum_{n=1}^{N_{c2}} \pi^{s,T_s}_{k,n} r^{s,T_s}_{k,n} \right). \tag{16}$$

In hybrid beamforming, the system hardware power consumption at eNB is also a significant parameter. Total power consumption at eNB includes total power dissipated predominantly in RF hardware circuitry and total power consumed in the transmission of bits. However, the system performances are constrained by the peak power expressed as

$$P_{peak}(S, N_{RF}) = \sum_{s \in \mathcal{S}_i} \sum_{k=1}^{K_s} \sum_{n=1}^{N_c} P_{k,n}^{s} + P_{nrf}(N_{RF}, P_t) \tag{17}$$

where $P_{nrf}(\cdot)$ is the power consumed at transmitter and $\mathcal{S}_i = \{s|\text{sectors concurrently served by } N_{RF} \text{ beams during the } i\text{th epoch}\}$. Another important parameter is the system fairness. In this paper, we use the geometric mean of long-run UE rate to to characterize fairness of long-run UE rates given as $\bar{G} = \sqrt[M]{\Pi_{s=1}^{S} \Pi_{k=1}^{K_s} R_k^s}$.

A. Resource allocation and cell sweeping schemes

We first discuss the problem of resource allocation in a sector, at a given time, implemented at baseband level of each RF unit. Here we have dropped the superscript τ/T_s for notation simplicity. To achieve a trade-off between system throughput and UE fairness within a sector at a time, we use proportional fairness (PF) subchannel allocation scheme. Accordingly, the optimization problem for resource allocation in sth active sector in an epoch is formulated as

$$\begin{aligned} (\mathcal{P}0): \quad & \max_{\pi_{k,n}^{s}, P_{k,n}^{s}} \sum_{k=1}^{K_s} \ln \left(\sum_{n=1}^{N_c} \pi_{k,n}^{s} r_{k,n}^{s} \right) \\ \text{s.t.} \quad & C01: \sum_{k=1}^{K_s} \pi_{k,n}^{s} \leq 1, \ \forall n; \ C02: \pi_{k,n}^{s} \in \{0,1\}, \ \forall k, n, s \\ & C03: P_{k,n}^{s} \geq 0, \ \forall k, n, s; \ C04: \sum_{k} \sum_{n} P_{k,n}^{s} \leq P_t, \ \forall s. \end{aligned} \tag{18}$$

Here, $\pi_{k,n}^{s}$ and $P_{k,n}^{s}$ are obtained by joint iteration and the complexity of such iterative method is very high to adopt in practical applications [34]. Therefore, we use a heuristic subchannel and power allocation approach as discussed next.

1) Subchannel and power optimization: In this approach we allocate ${\pi_{k,n}^{s}}^*$ and ${P_{k,n}^{s}}^*$ using the fact that system performance evaluation in our work is a relative analysis of rates over all the sectors. To ensure proportional fairness condition the optimization problem for subchannel allocation in sth sector served during an epoch is formulated as

$$(\mathcal{P}1): \quad \max_{\pi^s_{k,n}} \sum_{k=1}^{K_s} \ln\left(\sum_{n=1}^{N_c} \pi^s_{k,n} r^s_{k,n}\right)$$
$$\text{s.t.} \qquad C11: \sum_{k=1}^{K_s} \pi^s_{k,n} \leq 1,\ \forall n;\ C12: \pi^s_{k,n} \in \{0,1\},\ \forall k,n \tag{19}$$

We relax $\pi^s_{k,n}$ to a real number in the interval $[0,1]$. Thus, Lagrangian formulation of $\mathcal{P}1$ is

$$\mathcal{L}(\pi^s_{k,n}, P^s_{k,n}, \lambda) = \sum_{k=1}^{K_s} \ln\left(\sum_{n=1}^{N_c} \pi^s_{k,n} r^s_{k,n}\right) - \sum_{n=1}^{N_c} \lambda_n \left(\sum_{k=1}^{K_s} \pi^s_{k,n} - 1\right) \tag{20}$$

where $\lambda = [\lambda_1, \cdots, \lambda_n]$ is the non-negative Lagrangian multiplier for constraint $C11$. By Karush-Kuhn-Tucker (KKT) condition we have

$$\frac{\partial \mathcal{L}(\pi^s_{k,n}, \lambda, \eta)}{\partial \pi^s_{k,n}} = \frac{r^s_{k,n}}{\sum_{n=1}^{N_c} \pi^s_{k,n} r^s_{k,n}} - \lambda_n \leq 0 \tag{21}$$

Denote $R^s_k = \sum_{n=1}^{N_c} \pi^s_{k,n} r^s_{k,n}$. In (21), if the nth subchannel is not allocated to kth UE then $\pi^s_{k,n} = 0, r^s_{k,n} = 0$ and $r^s_{k,n}/R^s_k - \lambda_n \leq 0$. On the other hand, if the nth subchannel is allocated to kth UE then $\pi^s_{k,n} = 1, r^s_{k,n} \neq 0$, and $r^s_{k,n}/R^s_k - \lambda_n > 0$. This implies that nth subchannel is allocated to kth by the following rule

$$k^* = \arg\max_k\ r^s_{k,n}/R^s_k. \tag{22}$$

To satisfy the constraint $C12$, λ_n is set between $r^s_{k,n}/R^s_k$ and one. Next, each UE is given a share of power equal to the number of of subchannels allocated to it. Thus, the optimization problem for power allocation for kth UE is formulated as

$$(\mathcal{P}2): \quad \max_{P^s_{k,n}} \sum_{n\in\omega_k} r^s_{k,n} \text{ s.t. } C21: P^s_{k,n} \geq 0,\ \forall k,n;\ C22: \sum_{n\in\omega_k} P^s_{k,n} \leq P_t|\omega_k|/N_c,\ \forall k \tag{23}$$

where ω_k is the set of subchannels allocated to kth UE. The Lagrangian of $\mathcal{P}2$ is formulated as

$$\mathcal{L}(P^s_{k,n}, \eta_k) = \sum_{n\in\omega_k} r^s_{k,n} - \eta_k \sum_{n\in\omega_k} \left(P^s_{k,n} - P_t|\omega_k|/N_c\right) \tag{24}$$

where η_k is the Lagrangian multiplier corresponding to $C22$. (24) is concave in $P^s_{k,n}$.

Hence, from KKT conditions we have

$$P^s_{k,n}{}^* = \left[\frac{(B/N_c)\pi^s_{k,n}{}^*}{\ln 2 \eta^*_k} - \frac{1}{\gamma^s_{k,n}}\right]^+ \text{ and } \eta^{(j+1)}_k = \left[\eta^{(j)}_k - \hat{\epsilon}^{(j)}\left(\frac{P_t|\omega_k|}{N_c} - \sum_{n\in\omega_k} P^s_{k,n}{}^*\right)\right]^+. \tag{25}$$

Here, η^*_k is the optimal dual variable for power constraint, which is updated using sub-gradient method with step size $\hat{\epsilon}$. The steps to solve for $\{\pi^s_{k,n}{}^*\}$ and $\{P^s_{k,n}{}^*\}$ are presented in Algorithm 1. For nth subchannel allocation, $\omega_k\ \forall k$ is calculated by assuming $\pi^s_{k,n} = 1\ \forall k$. Then $P^s_{k,n}$ and $r^s_{k,n}\ \forall k,n$ are found, and $R^s_k \leftarrow \sum_{n\in\omega_k} r^s_{k,n}$. Accordingly, optimal k^* is evaluated using (22). Thereafter $P^s_{k,n}$ and R^s_k are again updated.

Algorithm 1 Proportional subchannel and power allocation per sector

Input: $\gamma_{k,n}^{s}, K_s$
Output: $\{\pi_{k,n}^{s}{}^{*}\}, \{P_{k,n}^{s}{}^{*}\}$
Initialize $\omega_k = 0$ and $R_k^s = 0\ \forall k$
for subchannels $n = 1$ to N_c **do**
    for users $k = 1$ to K_s **do**
        Set $\pi_{k,n}^{s} = 1$ and $\pi_{l,n,s} = 0$ for $l \neq k$
        Find ω_k and $\{P_{k,n}^{s}\}$ using (25)
        Update $\{r_{k,n}^{s}\}$ and $R_k^{s\prime} = R_k^s + r_{k,n}^s$
    end for
    Find k^* with $R_k^s \leftarrow R_k^{s\prime}$ in (22)
    Update $\pi_{k,n}^{s}{}^{*}$, $P_{k,n}^{s}{}^{*}$ and R_k^s
end for

2) Cell sweeping schemes: In the proposed variable frame structure, if sth sector served by bth beam during ith epoch has $K_s = 0$ then the frame duration is τ, and if $K_s \neq 0$ then the frame duration is $\tau + T_s$. Further, since zero-forcing or MMSE precoder is not used at the baseband, another way to reduce inter-beam interference while increasing net throughput is to optimally allocate the subchannel and power to the UEs in the concurrently active beams. Also, the inter-beam interference reduction is possible only when the beams move in synchronized manner. Hence, with N_{RF} concurrent beams, two cell sweeping schemes, namely, synchronous and asynchronous cell sweeping are possible.

During the ith epoch of synchronous cell sweep, the sojourn time of sectors served by the concurrent beams is the same, i.e., $t_s^i = \tau + T_s \triangleq \tau + T\ \forall s \in \mathcal{S}_i$, where $\mathcal{S}_i = \{s|\text{sectors concurrently served by } N_{RF} \text{ beams during the } i\text{th epoch}\}$. If the sth sector served by bth beam has $K_s = 0$, the beam b remains silent during t_s^i (that is, it does not transmit, saving power) while the other concurrently active beams service their respective sectors. Note that, sector sojourn time is not necessarily equal to frame duration. Using synchronous cell sweep, it is possible to achieve inter-beam interference reduction. On the other hand, in asynchronous cell sweep, each beam has the freedom to move independently over its set of sectors $\mathcal{S}_b$ without wasting time, i.e., beam $b' \neq b$ can move to another sector $s \in \mathcal{S}_{b'}$ while the beam b continues to serve its currently scheduled sector. Although by using asynchronous cell sweeping coordinated resource allocation over the concurrently active sectors is not possible, resources can be optimized per

Algorithm 2 Resource allocation in synchronous cell sweep scheme

Input: S, N_{RF}, and total number of epochs I
Output: $\{\pi^s_{k,n}{}^*\}, \{P^s_{k,n}{}^*\}\ \forall k, n, s$
for epoch $= 1$ to I **do**
    Find the set of active sectors $\mathcal{S}_i = \{\forall s_b | \text{sectors being served at } i\text{th epoch with } K_{s_b} > 0\}$
    Initially assume zero inter-beam interference
    for $s \in \mathcal{S}_i$ **do**
        Find $\{\pi^s_{k,n}\}, \{P^s_{k,n}\}\ \forall n, k$ using Algorithm 1
    end for
    Update $\gamma^s_{k,n}\ \forall s \in \mathcal{S}_i$ according to (8) and $R^s_k\ \forall k$
    Go to step 6 till $\sum_{s\in\mathcal{S}_i}\sum_{k=1}^{K_s} R^s_k$ converges
end for

beam at a time to improve the beam throughput. This results in increased inter-beam interference compared to the synchronous cell sweeping scheme. Nevertheless, it facilitates in improving the fairness of long-run UE rates by optimizing the sector sojourn times. The steps for resource allocation in synchronous and asynchronous cell sweep schemes are presented in Algorithm 2 and 3, respectively. When even one of the N_{RF} beams steers, it is counted as a new epoch in the algorithms. Also, OFDM resource allocation is performed separately for τ and T duration of the frame consisting of N_{c1} and N_{c2} subchannels, respectively.

Remark 2. *The optimal value of sector sojourn time t^i_s of sth sector served by bth beam during*

Algorithm 3 Resource allocation in asynchronous cell sweep scheme

Input: S, N_{RF}, and total number of epochs I'
Output: $\{\pi^s_{k,n}{}^*\}, \{P^s_{k,n}{}^*\}\ \forall k, n, s$
Initialize the set of active sectors $\mathcal{S}_0 = \{\forall s_b | \text{sectors steered at } i\text{th epoch with } K_{s_b} > 0\} = \emptyset$
for $i = 1$ to I' **do**
    Initially assume zero inter-beam interference,
    Find $\mathcal{S}_i$ and the set of sectors steered $\mathcal{S}^{steer}_i = \{s \in \mathcal{S}_i | s \in \mathcal{S}_i / \mathcal{S}_{i-1}\}$
    for $s \in \mathcal{S}^{steer}_i$ **do**
        Find $\{\pi^s_{k,n}\}, \{P^s_{k,n}\}\ \forall n, k$ using Algorithm 1
        Update $\gamma^s_{k,n}\ \forall s \in \mathcal{S}^{steer}_i$ according to (8)
    end for
end for

ith epoch, given total cell sweep time T_{tot}, for the two cell sweeping schemes are found as

$$\begin{aligned} &\textit{Synchronous: } t_s^i = \tau + \Big\lfloor (T_{tot} - |\mathcal{S}_b|\tau)/(T_{OFDM}|\mathcal{S}_b|) \Big\rfloor T_{OFDM} = \tau + T \quad \forall i, s, b \\ &\textit{Asynchronous: } t_s^i = \begin{cases} \tau, & K_s = 0, \\ \tau + T_s, & K_s \neq 0 \end{cases} \end{aligned} \tag{26}$$

where T_s is found to optimize system performance.

B. Joint estimation of number of sectors and concurrent beams

This section specifies the problem formulation for joint estimation of the optimal number of concurrent beams N_{RF}^* and optimal number of sectors S^* (or sector beamwidth Θ^*) required for average long-run UE rate maximization for a given eNB power budget.

From (8) we observe that, on one hand increasing N_{RF} will increase side-lobes interference, thereby decreasing cell peak data rate of UE in an epoch, while on the other hand it increases sector sojourn time. Thus, we need to choose a suitable value of N_{RF} that maximizes the average long-run UE rate $\bar{R}$. Moreover, S also influences the amount of inter-beam interference and sector sojourn time. Further, (17) indicates that peak power is also a function of N_{RF} and S (since P_t is a function of S). Therefore, the optimization problem to estimate S^* and N_{RF}^* that achieves highest $\bar{R}$ for a given eNB power budget P_{budget} is formulated as

$$\begin{aligned} (\mathcal{P}3): \quad & \max_{\pi_{k,n}^s, P_{k,n}^s, S, N_{RF}} \bar{R}(\pi_{k,n}^s, P_{k,n}^s, S, N_{RF}) \\ \text{s.t.} \quad & C31: \sum_{k=1}^{K_s} \pi_{k,n}^s \leq 1, \ \forall n, s; \ C32: \pi_{k,n}^s \in \{0,1\}, \ \forall k, n, s \\ & C33: \sum_{n=1}^{N_c} \sum_{k=1}^{K_s} P_{k,n}^s \leq P_t, \ \forall s; \ C34: P_{k,n}^s \geq 0, \ \forall k, n, s \\ & C35: S \in \mathcal{S}; \ C36: S \geq N_{RF}; \ C37: 0 \leq P_{peak} \leq P_{budget}. \end{aligned} \tag{27}$$

However, $\mathcal{P}3$ is jointly non-convex in $(\pi_{k,n}^s, P_{k,n}^s, S, N_{RF})$ and is a NP-hard problem. Therefore, we decompose $\mathcal{P}3$ and decouple the optimal resource allocation of UEs in all the sectors and

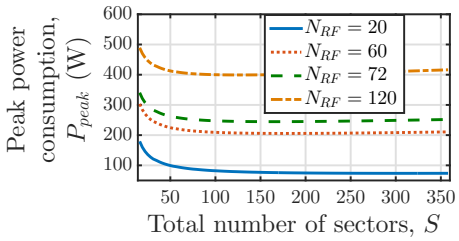


Figure 4: P_{peak} versus sectorization scheme for different N_{RF}.

optimal (S^*, N_{RF}^*) search. Constraint $C36$ has a straightforward implication that the total number of beams cannot exceed the total number of sectors in the cell. It is notable that water filling allocates power over subchannels so as to maximize sum rate. Consequently, the difference between the total available power P_t and total allocated power $\sum_{k=1}^{K_s}\sum_{n=1}^{N_c} P_{k,n}^s$ is nearly zero. As a result, from (17) we have $P_{peak} = N_{RF}P_t + P_{nrf}(N_{RF}, P_t)$ i.e., $C37$ is relaxed to $0 \leq N_{RF}P_t + P_{nrf}(N_{RF}, P_t) \leq P_{budget}$. Therefore, the feasible value of N_{RF} is limited by the available power budget P_{budget}, as shown in Fig. 4. Accordingly, we fix the value of N_{RF} and set S, and find the corresponding transmit power P_t given by (6). Thereafter, we use Algorithm 2 or 3 for resource allocation in all the epochs. $\bar{R}$ is computed by substituting value of t_s^i in (16) as specified in Remark 2. The steps of the algorithm are repeated by incrementing N_{RF} to search for S^* while satisfying $C35$ and $C36$.

C. Optimal sector sojourn time estimation for non-homogeneous UE distribution

In this section, we consider M non-homogeneously distributed UEs within the cell while the rest of the system parameters remain unchanged. Let $t_s = \tau + T_s$ be the sector sojourn time of a sector s. It is notable that, in case of homogeneous UE distribution, $T_s = T$ given by (26) was optimal. However, when the variance of UEs distribution among the sectors is high, the round-robin scheduling of sectors results in UE starvation long-run basis in densely populated sectors. Finding the best sectorization schemes for each beam individually is one way to increase the fairness of long-run UE rates for a given value of N_{RF}^*. This method requires a search of $|\mathcal{S}|^{N_{RF}}$ combinations. Another way to improve fairness is to optimize t_s. Therefore, we propose WRR scheduling in which t_s is calculated to maximize the fairness of long-run UE rates.

Here, we present the analysis for a single beam b out of parallel N_{RF} beams, serving $\mathcal{S}_b$ sectors in a cell sweep duration T_{tot}. The same analysis is applicable for the rest of the beams. The fairness index $\bar{G}$ is the highest when all the UEs have the same R_k^s $\forall k, s \in \mathcal{S}_b$. In a sectored cell system model, however, optimizing the scheduling time for each UE is not feasible. Nevertheless, we characterize the fairness in terms of expected long-run UE rate within a sector, $\bar{R}_s = \sum_{k=1}^{K_s} R_k^s/K_s$. Hence, as K_s increases the fraction of resources per UE in a sector decreases. Therefore, different sectors will experience different mean UE rate. For equal fairness guarantee, let $\bar{R}_s t_s = C$ (constant) $\forall s \in \mathcal{S}_b$ such that $\sum_{s\in\mathcal{S}_b} t_s \leq T_{tot}$, assuming $t_s \in \mathbb{R}^+$. Thus,

$$\sum_s t_s = \sum_s \frac{C}{1/\bar{R}_s} \Rightarrow C = \frac{T_{tot}}{\sum_s 1/\bar{R}_s} \Rightarrow t_s = \frac{T_{tot}/\bar{R}_s}{\sum_{s'}/\bar{R}_{s'}}. \tag{28}$$

Algorithm 4 Integer LP to find optimal sector sojourn time

Input: $\bar{R}_s \ \forall s \in \mathcal{S}_b$, T_{tot}, T_{OFDM}
Output: $\mathbf{t}^* = \{n_s^*\} \ \forall s \in \mathcal{S}_b$
Solve for $\mathbf{t}^0 = \{n_s\}$ using LP in (28) with integer restriction relaxed
Initialize: $j = 0; f_{max} = f(\mathbf{t}^0); \mathbf{t} = \emptyset$
while $\mathcal{S}_b \neq \emptyset$ **do**
    $s_j = \mathcal{S}_b\{1\}, j = j + 1$
    $\mathcal{S}_b = \mathcal{S}_b \backslash \{s_j\}$
    Set $n_j^+ = \lceil \mathbf{t}^0(j) \rceil$ and $n_j^- = \lfloor \mathbf{t}^0(j) \rfloor$
    Find $\hat{\mathbf{t}}^+ = \{n_s \forall s \in \mathcal{S}_b | n_1, \cdots, n_{j-1}, n_j^+\}$; $\hat{\mathbf{t}}^- = \{n_s \forall s \in \mathcal{S}_b | n_1, \cdots, n_{j-1}, n_j^-\}$ from (28)
    substituting $T_{tot} = T_{tot} - |\mathcal{S}_b|\tau - \sum_{i=1}^{j} n_s T_{OFDM}$
    if $f(\hat{\mathbf{t}}^+) \leq f_{max}$ and $f(\hat{\mathbf{t}}^+) \geq f(\hat{\mathbf{t}}^-)$ **then**
        $\mathbf{t}^*(j) = n_j^+$
    else
        $\mathbf{t}^*(j) = n_j^-$
    end if
end

Accordingly, the weights for determining t_s in WRR are $w_s = 1/\bar{R}_s$. However, t_s should be integer multiple of T_{OFDM}. Therefore, the optimization problem to optimize sector sojourn time (excluding beacon subframe in each sector) is expressed as

$$\begin{aligned} (\mathcal{P}4): \quad & T_s = \frac{w_s}{\sum_{s' \in \mathcal{S}_b} w_{s'}} (T_{tot} - |\mathcal{S}_b|\tau) \triangleq n_s T_{OFDM} \\ \text{s.t.} \quad & C41: n_s \in \mathbb{Z}^+, \ \forall s; \ C42: \sum_{s \in \mathcal{S}_b} n_s T_{OFDM} \leq T_{tot} - |\mathcal{S}_b|\tau. \end{aligned} \tag{29}$$

The problem $\mathcal{P}4$ is an integer linear programming (LP) optimization problem and solution to the problem is found using Algorithm 4.

Remark 3. *Let $f(\mathbf{t}) = \sum_{s \in \mathcal{S}_b} \bar{R}_s n_s T_{OFDM}$ be the total cell throughput, where $\mathbf{t} = \{n_s\}_{s \in \mathcal{S}_b}$. The solution $\mathbf{t} = \mathbf{t}^0 \in \mathbb{R}^+$ obtained from* (28) *gives an upper bound $f_{max} = f(\mathbf{t}^0)$. Let $\mathcal{T}$ be the feasible search region of $\mathbf{t}$ satisfying $C41$ and $C42$. For some $\mathbf{t}' \in \mathcal{T}$, the condition $f(\mathbf{t}') > f(\mathbf{t}^0)$ is possible only if sectors with higher $\bar{R}_s$ (equivalently smaller K_s) are allocated more time, hence contradicting the required objective of WRR. Therefore f_{max} acts as the upper bound on the throughput maximization while solving for $\mathcal{P}4$.*

Consequently, the LP reduces to the form

$$(\mathcal{P}5): \underset{\{n_s\}\in\mathbb{Z}^+}{\text{argmax}}\ f(\mathbf{t}) = \sum_{s\in\mathcal{S}_b} \bar{R}_s n_s T_{OFDM}$$
$$\text{s.t.}\quad C51: \lfloor \mathbf{t}^0(s) \rfloor \leq n_s \leq \lceil \mathbf{t}^0(s) \rceil,\ \forall s;\ C52: \sum_{s\in\mathcal{S}_b} n_s T_{OFDM} \leq T_{tot} - |\mathcal{S}_b|\tau;\ C53: f(\mathbf{t}) \leq f(\mathbf{t}^0). \tag{30}$$

The steps to find optimal $\mathbf{t}^* = \{n_s^*\}_{s\in\mathcal{S}_b}$ are presented in Algorithm 4. The search for $\mathbf{t}^0 \in \mathbb{R}^+$ begins by finding w_s from (28). Subsequently, the objective is to maximize $f(\mathbf{t})$ while searching for $t \in \mathbb{Z}^+$. The constraint $C51$ reduces the search space considerably by bounding each n_s to two nearest integer values of $\mathbf{t}^0(s)$ only. The explanation of upper bound on n_s is given in Remark 3, and the lower bound on n_s is justified by the fact that setting $n_s < \lfloor \mathbf{t}^0(s) \rfloor$ will lower the throughput. In Algorithm 4, at jth iteration the jth sector $s_j = \mathcal{S}_b\{1\}$ is selected to find the corresponding optimal n_j^*, and $\mathcal{S}_b$ is updated by excluding s_j. The two possible integer values of n_j are found as $n_j^+ = \lceil \mathbf{t}^0(j) \rceil$ and $n_j^- = \lfloor \mathbf{t}^0(j) \rfloor$. Thereafter, $\hat{\mathbf{t}}^+(\hat{\mathbf{t}}^-) \in \mathbb{R}^+$ for all the remaining sectors $s \in \mathcal{S}_b$ is found from (28) given $\{n_1^*, \cdots, n_{j-1}^*\}$ and $n_j^+(n_j^-)$. If the solution is feasible and $f(\hat{\mathbf{t}}^+) \geq f(\hat{\mathbf{t}}^-)$, then $\mathbf{t}^*(j) = n_j^+$; otherwise $\mathbf{t}^*(j) = n_j^-$.

V. Complexity Analysis

Below, we present the computational complexity of different modules.

1) Subchannel and power allocation per sector: For a system with K_s UEs and N_c subchannels in a sector, $\mathcal{P}0$ has a complexity of $\mathcal{O}(\zeta_R(K_s N_c + \zeta_\eta(N_c^2 \log N_c)))$ [34], where ζ_R and ζ_η denote the number of iterations required for convergence of rate and power allocation, respectively. Using the heuristic subchannel and power optimization problem, the combined complexity of resource allocation of $\mathcal{P}1$ and $\mathcal{P}2$ is reduced to $\mathcal{O}(N_c \sum_{k=1}^{K_s} \eta_P |\omega_k| \log(|\omega_k|)) = \mathcal{O}_1$. Here, η_P is the number of iterations required for convergence of power allocation.

2) Cell sweeping schemes: Let ζ_R' be the number of iterations required for convergence of cell peak sum rate for coordinated interference reduction in Algorithm 2. Thus, Algorithm 2 has a complexity of $I\zeta_R' N_{RF} \mathcal{O}_1$ per epoch, where I is the number of epochs in synchronous cell sweeping. In Algorithm 3, on the other hand, due to non-coordinated resource allocation, the resource optimization step in the sectors is only performed once per epoch. Thus, the complexity of Algorithm 3 is reduced to $I' N_{RF} \mathcal{O}_1$, where I' is the total number of epochs in asynchronous cell sweeping. The complexity of Algorithm 2 is smaller as $I' \leq I\zeta_R'$. A worst-case comparison is performed over all the combinations of S and N_{RF} to jointly estimate S^* and N_{RF}^*.

3) Estimation of sector sojourn time: In Algorithm 4, after solving for $\mathtt{t} \in \mathbb{R}^+$ using (28), the integer optimal $\mathtt{t}^*$ is found sequentially for all the sectors covered by a beam, i.e., the sector sojourn time estimation loop is executed $|\mathcal{S}_b|$ per beam having a total complexity of $\mathcal{O}(N_{RF}|\mathcal{S}_b|)$.

VI. Results and Discussions

In this section, we present the numerical simulation results generated using MATLAB. The simulation set-up configurations are given in Table III. First, we present the simulations illustrating the advantages of the proposed variable time frame structure and then proceed using this MAC protocol as an underlying access mechanism for studying optimum system performance with homogeneous and non-homogeneous UE distribution scenarios.

The numerical results are compared with the 3GPP NR scheme as a baseline. For beam management and data transmission, the total time in the 3GPP NR scheme is divided into control and data parts [19]. The IA technique is explained in Section III-A and is performed during beam management phase. The UEs are grouped into N_{RF} sets based on AoD similarities in 3GPP NR scheme [10]. Furthermore, at most N_{RF} UEs are scheduled in time domain for data transmission with a dedicated RF unit per UE in a group employing baseband precoding to reduce IUI. Note that, as per the 3GPP NR specifications, a slot is of fixed size 14 OFDM symbols and a beam stays in a specific direction for an integer multiple of a slot duration.

Table III: Simulation parameters and values

Parameter	**Description**	**Value**	**Parameter**	**Description**	**Value**
D	Cell Diameter	400 m	f_c	Carrier frequency	28 GHz
B	Bandwidth	1 GHz	θ	HPBW	1^o to 25^o
U	Number of UEs	$100, 600$	N_c	Number of subchannels	32
d'	Inter-element ULA spacing	$\lambda_c/2$	S	Number of sectors	14 to 360
N_{RF}	Number of concurrent beams	$\leq S$	N_0	Noise spectral density	-174 dBm/Hz
K_R	Rician fading parameter	8 dB	T_{tot}	SS block periodicity	20 ms
$\delta\phi$	zone angular granularity	1^o	T_{tot}	SS burst periodicity	20 ms
EIRP	Effective isotropic radiated power	52 dBm	N_t	Number of antenna elements per RF unit at eNB	5 to 115
L_k	Number of MPCs	1 (LOS), 4 (NLOS)	$\alpha_{k,l}$	lth MPC fading parameter	Rician (LOS), Rayleigh (NLOS)
N_r	Number of antenna elements at the UE	4	S_r	Number of UE steerable directions	12

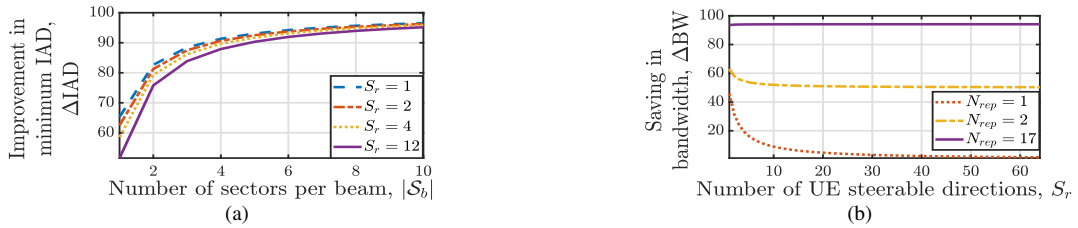


Figure 5: (a) Improvement in IAD with $N_{rep} = 17$ and (b) saving in bandwidth using proposed IA procedure over the 3GPP NR. $N_{rep} = 17$ isequivalent to using full 1 GHz bandwidth for SS blocks

A. Performance evaluation of the proposed frame structure

1) IAD: With the proposed frame structure, considering no contention, IA of a UE can be performed within a beacon duration, whereas 3GPP NR uses exhaustive beam search over the entire region before final RAR transmission to the UE. Here, we consider $t_{c2} = 5T_{OFDM}$ and $t_{c4} = 4T_{OFDM}$ (explanation given in Appendix C). Fig. 5(a) shows the percentage of improvement in IAD with the proposed IA procedure over 3GPP NR. Further, using the proposed IA procedure a considerable amount of BW (given by (14)) is saved, as shown in Fig. 5(b).

2) Multiple UEs per beam: Figs. 6(a) and 6(b) illustrate the advantage of serving multiple UEs at a time located in a sector by a beam using wideband mmWave channel. The decline in $\bar{R}$ with number of UEs per sector is non-linear. Thus, it is suitable to serve multiple UEs at a time over wideband mmWave channel. We compare the performances of the proposed frame structure with two other competitive standalone mmWave communication schemes: 3GPP NR scheme and perfect eNB-UE beam alignment with variable TTI frame structure scheme [21]. Figs. 6(a) and 6(b) show that the $\bar{R}$ of UEs and the energy efficiency of the eNB (ratio of total bits transmitted to total energy consumed) outperform the 3GPP NR model and are comparable to variable TTI with perfect beam alignment scheme. Nonetheless, it is not possible to achieve perfect beam alignment practically.

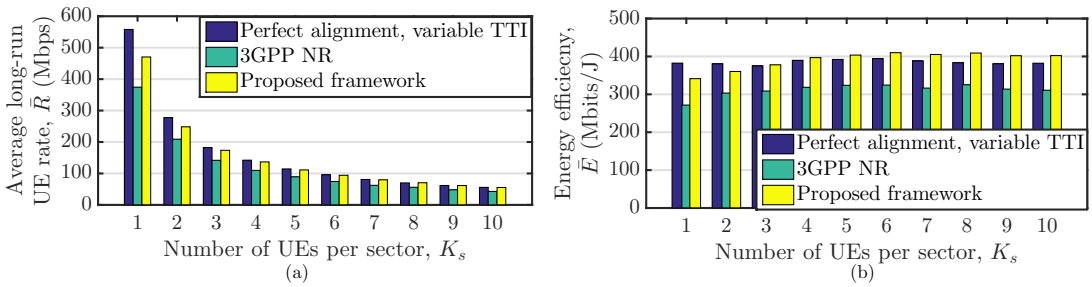


Figure 6: (a) Comparison of the (a) average long-run UE rate and (b) eNB energy efficiency achieved using with the existing protocols and proposed protocol. Here, $|\mathcal{S}_b| = 5$ and $S_r = 12$.

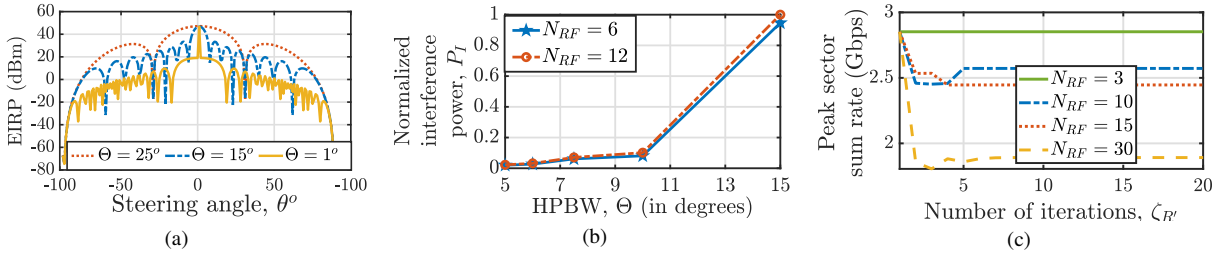


Figure 7: (a) EIRP level in main-lobe and side-lobes in a beam at different beamwidth. (b) Normalized interference experienced by a UE with respect to variable HPBW. (c) Convergence of peak sector sum rate in an epoch using Algorithm 2 for $S = 30$ and $M = 100$.

B. Optimum S^* and N^*_{RF} for homogeneous UE distribution

Fig. 7(a) shows the EIRP level in main-lobe and side-lobes for beams of different width attained from analog beamforming vector given by (7). Even though narrow beams have a high gain, modulating P_t for maintaining constant EIRP results in a reduced effective radiated power from the interfering side-lobes. Moreover, the cell coverage range is maintained constant, and the peak EIRP is the same irrespective of the beamwidth. Fig. 7(b) shows that the normalized interference power P_I, from side-lobes of interfering beams, experienced by a UE in a sector increases with increased Θ (equivalent to reduced S). It is notable, that power allocation using Algorithm 1 has $\eta_p < 100$ when $\eta_k^{(0)} = 1\ \forall k$, $\hat{\epsilon} = 0.05$ and convergence tolerance $= 10^{-4}$. We also observe that P_I increases with a higher value of N_{RF}. From Fig. 7(c) it is verified that solving for subchannel and power resource allocation in Algorithm 2, the peak sector sum rate in an epoch converges in less than 20 iterations; in 1^{st} iteration peak sum rate is maximum and invariant of N_{RF} due to the initial assumption of zero interference per subchannel. Further, from Fig. 7(c) we study that, for fixed S, increasing N_{RF} degrades the peak sector throughput, which is because of the increase in P_I. Note at $N_{RF} = 3$, the beams are 120^o apart from each other and have negligible inter-beam interference, thus the peak sum rate remains constant.

Remark 4. *Despite a decrease in peak throughput per sector, the average long-run UE rate*

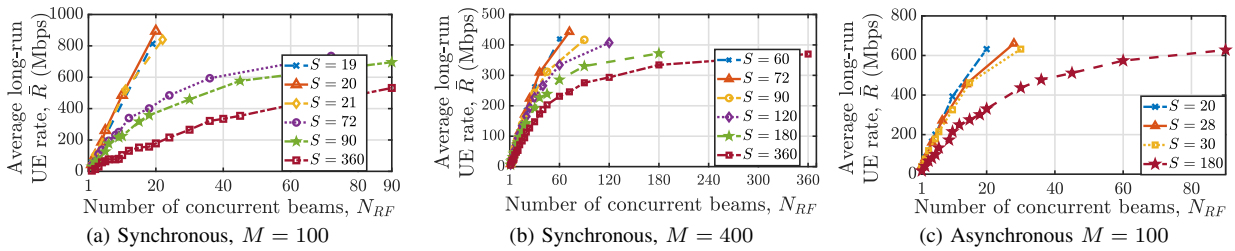


Figure 8: Illustration of $\bar{R}$ as a function of S and N_{RF}.

$\bar{R}$ improves with increased N_{RF} for a fixed S. The improvement in $\bar{R}$ with N_{RF} is because of the fact that the increase in sector sojourn time from using a large number of concurrent beams overcomes the rate loss due to increased interference level. Therefore, for a given S, the maximum $\bar{R}$ is achieved at $N_{RF} = S$.

From Fig. 8(a) and 8(b), we observe that even with $N_{RF} = S$, the highest value of $\bar{R}$ is achieved at $S^* = 20$ and $S^* = 72$, respectively, for $M = 100$ and $M = 400$ homogeneously distributed UEs. This is so because, for $S < S^*$ (or equivalently $\Theta > \Theta^*$) the interference from sidelobes of concurrent beams is high, as noted from Fig. 7(a). For $S > S^*$ even though effective interference from sidelobes of a single concurrent beam is small, because of a large value of N_{RF} the cumulative interference is increased which degrades $\bar{R}$. Hence, using the narrowest possible beam is not beneficial. Results in Figs. 8(a) and 8(b) are generated using synchronous cell sweeping. From Fig. 8(c) we observe that for asynchronous cell sweeping at $M = 100$ the value of $S^* = 28$ which is slightly higher as compared to synchronous cell sweeping ($S^* = 20$) although maximum attainable $\bar{R}$ is a little reduced. Furthermore, P_{budget} limits the number of beams N_{RF} that can be activated at eNB, which influences S^* selection. For example, the optimal parameters are $(S^*, N^*_{RF}) = (30, 15)$, with $\bar{R} = 210.58$ Mbps and $P_{peak} = 49.07$ W, when $M = 400$ and the $P_{budget} = 50$ W (Fig. 4). If eNB has no peak power constraint, the best answer is $(S^*, N^*_{RF}) = (72, 72)$, with maximum $\bar{R} = 458.72$ Mbps and $P_{peak} = 271.63$ W.

Table IV compares the scheduling complexity of the proposed model over the existing competitive UE scheduling schemes, using fully and partially connected hybrid beamforming structure, as a function of M, N_{RF}, and N_t at the eNB. As mentioned in Section I-A, [9]–[11] dedicate one RF chain per UE at a time using 3GPP NR MAC as underlying protocol. In our proposed model all the UEs that are located within the beam coverage area are served simultaneously.

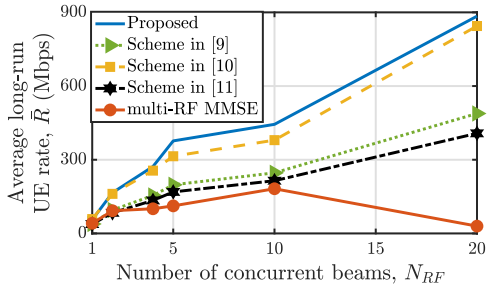


Figure 9: Comparison of achievable $\bar{R}$ of the proposed system model with the schemes in [9]–[11] and multi-RF MMSE method for $M = 100$.

Table IV: UE scheduling complexity

User scheduling Method	Hybrid structure	Complexity
Channel correlation [9]	Fully-connected	$\mathcal{O}(M)$
AoD's similarities [10]	Fully-connected	$\mathcal{O}(MN_{RF}\eta_I)$
Virtual beamforming [11]	Partially-connected	$\mathcal{O}(N^2_{RF})$
Proposed method	Partially-connected	$\mathcal{O}(N_{RF})$

η_I: number of iterations for algorithm convergence

Hence, scheduling is independent of M and depends only on the N_{RF}. Further, Fig. 9 shows that the proposed sectored-cell model using multi-analog beams with optimum beamwidth for serving multiple UEs per beam achieves better $\bar{R}$ and also with reduced scheduling complexity compared to the models in [9]–[11]. For the sake of fair comparison, we fix the total number of antenna elements at eNB equal to 256 and maintain $N_t \leq 256/N_{RF}$ in the proposed model.

We also analyze a scenario in which multi-RF beams in a sectored cell model serve multiple UEs per beam over OFDMA, where, instead of the proposed power allocation in Algorithm 2, the baseband utilizes an MMSE precoder per subchannel to minimize inter-beam interference. We denote this as multi-RF MMSE method. Besides high computation complexity, the multi-RF MMSE approach has smaller $\bar{R}$, as shown in Fig. 9, because the worst UE channel condition acts as a bottleneck to its performance. In contrast, the approaches in [9]–[11] only distribute subchannels to UEs with the best channel conditions and thus have better performance compared to the multi-RF MMSE approach. Though in multi-RF MMSE approach the spatial multiplexing gain is large at higher value of N_{RF}, $\bar{R}$ degrades after certain high value of N_{RF}. This is because in sectored-cell model, the RF beams are not perfectly orthogonal, and hence dominant eigenmodes are close to zero in some beam directions, thus limiting the throughput.

C. Optimal sector sojourn time estimation for non-homogeneous UE distribution using WRR

We set $Z = N_{RF}$ in (1) for non-homogeneous UE set-up such that $|\mathcal{S}_b| = 4$ for $S^* = 28$ at $M = 100$. Fig. 10(a) illustrates that with increasing Υ asynchronous scheme offers improved $\bar{R}$ as opposed to allocating equivalent sector sojourn time to all sectors in the synchronous scheme. Also, compared to the 3GPP NR scheme a notable gain in $\bar{R}$ with the asynchronous scheme is achieved because bandwidth consumed in control overhead is less using variable time frame at MAC layer (shown in Fig. 5(b)) and the wideband channel is efficiently utilized by serving

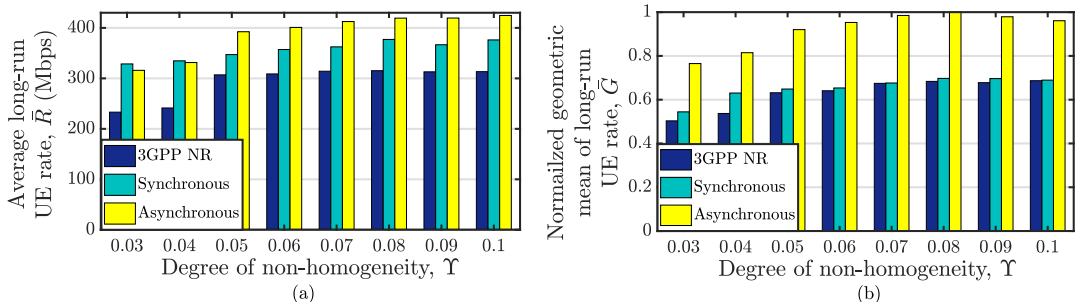


Figure 10: Comparison of (a) $\bar{R}$ and (b) normalized $\bar{G}$ for $M = 100$ non-homogeneously distributed UE achieved with 3GPP NR, synchronous and asynchronous schemes.

multiple UEs at a time. Moreover, significant improvement in fairness $\bar{G}$ of long-run UE rates with asynchronous over synchronous and 3GPP NR scheme is achieved, as shown in Fig. 10(b). This is so because the sectors with higher K_s (hence, smaller $\bar{R}_s$) are allocated more time.

VII. Conclusion

In this paper, we have investigated the outdoor mmWave scenario with a much larger UE population than the number of available RF chains. To support UEs in a cell divided into sectors, we used multi-RF chains. At any given time, each RF unit serves a single sector, with multiple UEs scheduled over the wideband channel using OFDMA. In comparison to other competing user grouping hybrid beamforming designs, the proposed model does not require brute zero-forcing precoding. From rate support findings it is observed that for a given cell sectorization, while interference from the side-lobes of concurrent beams decreases a UE's peak data rate, the time-averaged data rate increases with higher number of RF chains. Also, it has been shown that using the narrowest beamwidth is not optimal. Further, it has been observed that raising the number of concurrent beams in a cell not only causes increased side-lobe interference but also greatly increases the eNB power budget. Hence, the power budget plays a crucial role in deciding the optimum number of concurrent beams and corresponding sector beamwidth. We have also proposed a variable time frame structure that requires less IAD and increases bandwidth utilization. Further, we have shown that by using the variable frame, we can efficiently schedule non-homogeneously distributed UEs, thus improving the UE fairness index as well as the average long-run UE rate.

The possible future works include comparison of the efficiency of variable time frame structures for different 3GPP NR frame methodology and extending the proposed sectored cell system to the case of uniform and non-uniform amplitude planar arrays that would require optimizing antenna amplitude, eNB height, elevation HPBW, and sector scheduling with UE fairness.

Appendix

A. *Proof of* $\int_{\hat{\phi}} \mathcal{Q}(ar(\hat{\phi}))d\hat{\phi} = 1$

$$\begin{aligned}\int_{\hat{\phi}=0^o}^{360^o} \mathcal{Q}(ar(\hat{\phi}))d\hat{\phi} &= \int_{\hat{\phi}=0^o}^{360^o} \frac{\exp(\Upsilon \lfloor \frac{\text{mod}(\hat{\phi}, 360^o/Z)}{\delta\phi} \rfloor \delta\phi)}{Z\sum_{i=0}^{(360^o/Z\delta\phi)-1} \exp(\Upsilon i\delta\phi)} d\hat{\phi} \\ &= \int_{\hat{\phi}=0^o}^{\frac{360^o}{Z}} I(\hat{\phi})d\hat{\phi} + \cdots + \int_{\hat{\phi}=\frac{(z-1)360^o}{Z}}^{\frac{z360^o}{Z}} I(\hat{\phi})d\hat{\phi} + \cdots + \int_{\hat{\phi}=\frac{(Z-1)360^o}{Z}}^{360^o} I(\hat{\phi})d\hat{\phi}\end{aligned} \tag{A.1}$$

where $I(\hat{\phi}) = \exp(\Upsilon \lfloor \frac{\text{mod}(\hat{\phi}, 360^o/Z)}{\delta\phi} \rfloor \delta\phi) / \left(Z \sum_{i=0}^{(360^o/Z\delta\phi)-1} \exp(\Upsilon i \delta\phi) \right)$ and $z = \{1, 2, \cdots, Z\}$. Let $\tilde{\phi} = \text{mod}(\hat{\phi}, 360^o/Z)$. Then (A.1) can be written as

$$\int_{\hat{\phi}=0^o}^{360^o} \mathcal{Q}(ar(\hat{\phi})) = Z \int_{\tilde{\phi}=0^o}^{360^o/Z} I(\tilde{\phi}) d\tilde{\phi} = Z \int_{\tilde{\phi}=0^o}^{360^o/Z} \frac{\exp(\Upsilon \lfloor \tilde{\phi}/\delta\phi \rfloor \delta\phi)}{Z \sum_{i=0}^{(360^o/Z\delta\phi)-1} \exp(\Upsilon i \delta\phi)} d\tilde{\phi}. \quad \text{(A.2)}$$

As per the definition, $\mathcal{Q}(ar(\hat{\phi}))$ is constant within the angular granularity of $\delta\phi$. Therefore, integration is replaced by summation and $\lfloor \tilde{\phi}/\delta\phi \rfloor \triangleq \tilde{\phi}/\delta\phi = j$, and hence

$$\int_{\hat{\phi}=0^o}^{360^o} \mathcal{Q}(ar(\hat{\phi})) = Z \sum_{j=0}^{(360^o/Z)/\delta\phi} \frac{\exp(\Upsilon j \delta\phi)}{Z \sum_{i=0}^{(360^o/Z\delta\phi)-1} \exp(\Upsilon i \delta\phi)} = 1. \quad \text{(A.3)}$$

B. Proof of Lemma 1

Lemma 1: When N_{RF} is sufficiently high, the probability of two or more UEs transmitting the same RACH preamble over PRACH in a single frame is very small.

Proof: Suppose the new UEs arrive according to Poisson distribution with mean arrival rate μ UEs/s/km^2 during a cell sweep duration, T_{tot}. With N_{RF} beams serving the cell, coverage area per beam is $Ar(N_{RF}) = \pi(D/2)^2/N_{RF}$. Then the probability of arrival of m new UEs during T_{tot} time interval is $\text{Pr}_a(m) = (\mu T_{tot} Ar(N_{RF}))^m e^{-\mu T_{tot} Ar(N_{RF})}/m!$. Consider that each new UE chooses a RACH preamble randomly from a set of $M_R = 64$ orthogonal preambles with equal probability. Then the probability that $k \geq 2$ out of m UEs chooses same RACH preamble is

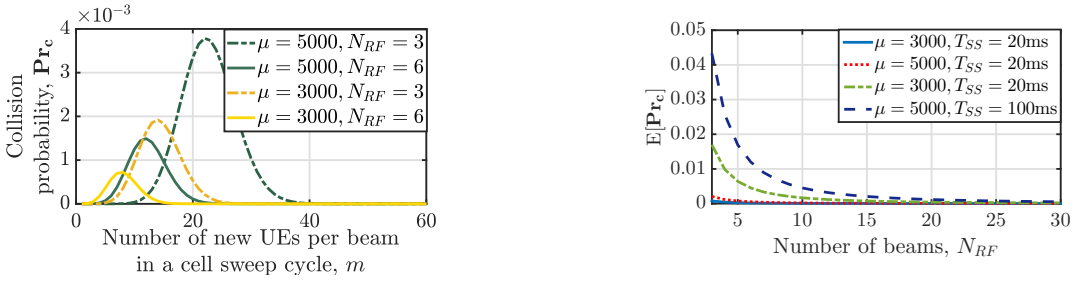
$$\text{Pr}_r(k \geq 2|m) = 1 - \binom{m}{0} \left(\frac{1}{M_R}\right)^0 \left(1 - \frac{1}{M_R}\right)^m - \binom{m}{1} \left(\frac{1}{M_R}\right) \left(1 - \frac{1}{M_R}\right)^{m-1}. \quad \text{(A.4)}$$

Thus, the probability of collision, Pr_c is obtained as

$$\text{Pr}_c = \text{Pr}_a(m) \times \text{Pr}_r(k \geq 2|m)$$

$$\mathbb{E}[\text{Pr}_c] = \sum_{m=2}^{\infty} \frac{(\mu T_{tot} Ar(N_{RF}))^m e^{-\mu\delta Ar(N_{RF})}}{m!} \left[1 - \frac{m}{M_R} \left(1 - \frac{1}{M_R}\right)^{m-1} - \left(1 - \frac{1}{M_R}\right)^m \right] \quad \text{(A.5)}$$

$$= \sum_{m=0}^{\infty} \frac{(\mu T_{tot} Ar(N_{RF}))^m e^{-\mu\delta Ar(N_{RF})}}{m!} \left[1 - \left(1 - \frac{1}{M_R}\right)^m \left(\frac{m}{M_R - 1} + 1 \right) \right].$$



(a) (b)

Figure A.1: (a) Pr_c and (b) $\mathbb{E}[\text{Pr}_c]$ of new UEs arriving during $T_{tot} = 100$ ms.

Here, we consider 3-sector LTE model as a reference. Fig. A.1(a) shows that maximum Pr_c is very small even for worst case scenario of $T_{tot} = 100$ ms, $N_{RF} = 3$, and μ as high as 5000 UEs/Km^2/s. Fig. A.1(b) shows that the expected collision probability decreases sharply on increasing N_{RF}, and in this set-up, $N_{RF} \geq 6$ can be considered a sufficiently high value.

C. Processing time calculation

The processing time is given as $t_{process} = 0.5N_I(2048 + 144)\kappa 2^{-\hat{\mu}}T_C$, where κ is a constant set to value 64, $\hat{\mu}$ corresponds to subcarrier spacing, i.e., $\Delta f = 240$ kHz $= 2^{\hat{\mu}} \times 15$ kHz, N_I is subcarrier-spacing dependent timing advance parameter, and $T_C = 1/(f_{max}N_f)$ [24]. The values $f_{max} = 480$ kHz and $N_f = 4096$ are fixed in 5G. Therefore, at $\Delta f = 240$ kHz, $t_{process} = N_I \times 2.23 \times 10^{-6}$ s. N_I is obtained from trip time, i.e., $2 \times$ propagation delay$/T_{OFDM}$, which is equivalent to 1 OFDM symbol, resulting in $t_{process} = 2.23\mu$s. Thus, the total time between the last SS block transmission and the start of PRACH is equal to the number of the transmission and processing times, i.e. $t_{c2} = 4T_{OFDM} + t_{process} \geq 5$ OFDM symbols. The same processing delay is applicable after PRACH transmission. Hence, $t_{c4} \geq T_{OFDM} + t_{process} = 2$ OFDM symbols.

REFERENCES

[1] N. Varshney and S. De, "Joint beamwidth and number of concurrent beams estimation in downlink mmWave communications," in *Proc. Nat. Conf. Commun.*, 2021, Kanpur, India, pp. 1–6.

[2] G. Sanfilippo, O. Galinina, S. Andreev, S. Pizzi, and G. Araniti, "A concise review of 5G new radio capabilities for directional access at mmWave frequencies," *Internet of Things, Smart Spaces, Next Generation Netw. Syst.*, pp. 340–354, 2018.

[3] T. S. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. N. Wong, J. K. Schulz, M. Samimi, and F. Gutierrez, "Millimeter wave mobile communications for 5G cellular: It will work!" *IEEE Access*, vol. 1, pp. 335–349, 2013.

[4] T. S. Rappaport, F. Gutierrez, E. Ben-Dor, J. N. Murdock, Y. Qiao, and J. I. Tamir, "Broadband millimeter-wave propagation measurements and models using adaptive-beam antennas for outdoor urban cellular communications," *IEEE Trans. Antennas and Propag.*, vol. 61, no. 4, pp. 1850–1859, 2012.

[5] F. Sohrabi and W. Yu, "Hybrid analog and digital beamforming for mmWave OFDM large-scale antenna arrays," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 7, pp. 1432–1443, 2017.

[6] B. Wang, M. Jian, F. Gao, G. Y. Li, and H. Lin, "Beam squint and channel estimation for wideband mmWave massive MIMO-OFDM systems," *IEEE Trans. Sig. Process.*, vol. 67, no. 23, pp. 5893–5908, 2019.

[7] G. Kwon and H. Park, "Limited feedback hybrid beamforming for multi-mode transmission in wideband millimeter wave channel," *IEEE Trans. Wireless Commun.*, vol. 19, no. 6, pp. 4008–4022, 2020.

[8] L. Qianrui, "Hybrid precoding for wideband multi-user MIMO millimeter wave system," in *Proc. IEEE Wireless Commun. Netw. Conf.*, 2019, pp. 1–6.

[9] A. Adhikary, E. Al Safadi, M. K. Samimi, R. Wang, G. Caire, T. S. Rappaport, and A. F. Molisch, "Joint spatial division and multiplexing for mm-wave channels," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1239–1255, 2014.

[10] A. Masmoudi and T. Le-Ngoc, "User grouping and hybrid RF/baseband precoding for multi-user massive MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 69, no. 10, pp. 11 308–11 322, 2020.

[11] E. Kim, J. Kwak, and S. Chong, "Virtual beamforming and user scheduling for sub-array architecture in mmWave networks," *IEEE Commun. Lett.*, vol. 23, no. 1, pp. 168–171, 2018.

[12] G. Kwon, N. Kim, and H. Park, "Millimeter wave SDMA with limited feedback: RF-only beamforming can outperform hybrid beamforming," *IEEE Trans. Veh. Technol.*, vol. 68, no. 2, pp. 1534–1548, 2018.

[13] V. N. Ha, D. H. Nguyen, and J.-F. Frigon, "Subchannel allocation and hybrid precoding in mmWave OFDMA systems," *IEEE Trans. Wireless Commun.*, vol. 17, no. 9, pp. 5900–5914, 2018.

[14] A. Alkhateeb, Y.-H. Nam, M. S. Rahman, J. Zhang, and R. W. Heath, "Initial beam association in millimeter wave cellular systems: Analysis and design insights," *IEEE Trans. Wireless Commun.*, vol. 16, no. 5, pp. 2807–2821, 2017.

[15] J. Zhang, Y. Huang, C. Zhang, S. He, M. Xiao, and L. Yang, "Cooperative multi-subarray beam training in millimeter wave communication systems," in *Proc. IEEE Global Commun. Conf.*, 2017, pp. 1–6.

[16] H. Soleimani, R. Parada, S. Tomasin, and M. Zorzi, "Fast initial access for mmWave 5G systems with hybrid beamforming using online statistics learning," *IEEE Commun. Mag.*, vol. 57, no. 9, pp. 132–137, 2019.

[17] Y. Li, J. G. Andrews, F. Baccelli, T. D. Novlan, and C. J. Zhang, "Design and analysis of initial access in millimeter wave cellular networks," *IEEE Trans. Wireless Commun.*, vol. 16, no. 10, pp. 6409–6425, 2017.

[18] Z. Xiao, P. Xia, and X.-G. Xia, "Codebook design for millimeter-wave channel estimation with hybrid precoding structure," *IEEE Trans. Wireless Commun.*, vol. 16, no. 1, pp. 141–153, 2016.

[19] 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, 2018.

[20] 3GPP, "5G: NR Physical layer procedures for data–Rel. 15," 2018, TS 38.214 V15.3.0.

[21] S. Dutta, M. Mezzavilla, R. Ford, M. Zhang, S. Rangan, and M. Zorzi, "Frame structure design and analysis for millimeter wave cellular systems," *IEEE Trans. Wireless Commun.*, vol. 16, no. 3, pp. 1508–1522, 2017.

[22] I. S. Association *et al.*, "Part 11: Wireless LAN medium access control (MAC) and physical layer (PHY) specifications," *IEEE Std.*, vol. 802, 2012.

[23] B. Schultz, "802.11 ad–WLAN at 60 GHz–A technology introduction," *Rohde & Schwarz*, 2013.

[24] 3GPP, "Physical layer procedures for data – Rel. 15," 2019, TS 38.214 V15.5.0.

[25] N. Varshney and S. De, "Optimum downlink beamwidth estimation in mmWave communications," *IEEE Trans. Commun.*, vol. 69, no. 1, pp. 544–557, 2021.

[26] 3GPP, "Study on channel model for frequency spectrum above 6 GHz–Rel. 14," 2017, TR 38.900 V14.2.0.

[27] M. R. Akdeniz, Y. Liu, M. K. Samimi, S. Sun, S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter wave channel modeling and cellular capacity evaluation," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1164–1179, 2014.

[28] O. El Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1499–1513, 2014.

[29] C. A. Balanis, *Antenna Theory: Analysis and Design*. John wiley & sons, 1997.

[30] FCC. (2015) *Notice of Proposed Rulemaking*. FCC-15-138A1.

[31] L. N. Ribeiro, S. Schwarz, M. Rupp, and A. L. de Almeida, "Energy efficiency of mmWave massive MIMO precoding with low-resolution DACs," *IEEE J. Sel. Topics Sig. Process.*, vol. 12, no. 2, pp. 298–312, 2018.

[32] 3GPP, "Physical channels and modulation – Rel. 15," 2020, TR 38.211 V15.8.0.

[33] N. Atteya, S. maximov, and M. A. El-saidny, "A new era for enhanced mobile broadband (White paper)," Mediatek, 2018.

[34] L. Zhang, C. Jin, and W. Zhou, "Decomposition proportional fairness algorithm for multiuser OFDM systems," in *Proc. IEEE Int. Conf. Commun. Workshops*, 2008, pp. 21–25.