Optimum Downlink Beamwidth Estimation in mmWave Communications

Optimum Downlink Beamwidth Estimation in mmWave Communications

Nancy Varshney, and Swades De

***Abstract*—The abstract goes here.With increasing density of data-hungry devices per unit area, allocating single highly-directed beam per user in millimeter-wave communications is not practical. Therefore, the requirement is to serve multiple users over a single beam. Considering a single-cell scenario with a fixed number of users, this paper addresses the problem of selection of optimal beamwidth depending on user density and distribution. First, by considering fixed beam service time in each sector, optimal beamwidth is estimated using exhaustive search for average long-run user rate and base station energy efficiency maximization. Based on the results of the average long-run user rate maximization using an exhaustive search, another method of reduced complexity is proposed to find sub-optimal beamwidth. Subsequently, optimum beamwidth is estimated with user density dependent variable time scheduling in a sector, that offers improved performances over fixed time scheduling. An efficient algorithm on variable time scheduling is also provided. Finally, the effect of localization error on optimal beamwidth estimation is investigated. The numerical results show that using the narrowest beam does not necessarily result in achieving a better average long-run user rate. Further, localization error does not affect the selection of optimal beamwidth, however, user Quality-of-Service degrades.**

***Index Terms*—Millimeter-wave communication, analog beamforming, scheduling, optimal beamwidth, localization error**

I. Introduction

Millimeter-wave (mmWave) spectrum (30-300 GHz) is envisaged to be a key resource to meet the high data demands of future wireless communication because of huge bandwidth availability. A major limitation of mmWave communication is high attenuation associated with a shorter wavelength. However, having a shorter wavelength in turn facilitates placing a large number of antenna elements over a very small area [1]. Thus, the gain realized through antenna beamforming overcomes high attenuation loss. As the array size is increased, the beam gets narrower with reduced footprint and enhanced directivity. The reduced antenna footprint of a highly directional beam leverages the benefits of spatial diversity in mmWaves, thereby providing increased spectral capacity. Conventional multiple-input multiple-output (MIMO) systems [2], [3], also known as digital beamforming, use a dedicated radio frequency unit (RF chain) per antenna element. But, using a large number of power-intensive RF chains per antenna is practically not feasible in mmWaves. Nonetheless, a highly narrow beam of high directivity can be formed using analog beamforming, consisting of a large number of active antenna elements connected to a single RF chain. Consequently, the power consumption of analog beamforming is comparatively less. The limitation of analog beamforming is that only a single beam per RF chain can be generated, and spatial data stream multiplexing is not possible. To this end we observe that, recent technological advancements have made it possible to switch and steer the beam of phased antenna arrays within negligible time (order of a few nanoseconds), thus making it possible to schedule angularly-apart users equipments (UEs) in the time domain using a single narrow beam.

UE localization using global positioning system (GPS) in conventional cellular communication has an error range in meters. Such high error is not tolerable in mmWaves employing directional beams. In mmWaves there are various localization methods, such as ray tracing, exhaustive beam searching, and RF-based ranging methods. The ray tracing method requires detailed surrounding and multipath components information, whereas exhaustive searching entails a large beam training data. The complexities of both these methods are proportional to the number of antenna elements at the base station (BS) and the UE. Two most popular localization methods in mmWaves are Angle-of-Arrival (AOA) and Time-of-Arrival (TOA) [4]. Localization accuracy in mmWaves outdoor scenarios, with large antenna array at the BS, single antenna at the UE, and line-of-sight paths using AOA and TOA methods, were studied in [5], [6]. It was shown in [6] that use of extended Kalman filtering can achieve sub-meter accuracy even for mobile devices.

A. Background and motivation

To reduce power consumption and improve spectral efficiency of the mmWave communication system, hybrid beamforming has been proposed in the literature. Hybrid beamforming is a trade-off between analog beamforming and digital beamforming. With a hybrid beamformer having N_{RF} RF chains at BS, N_{RF} data streams can be generated simultaneously using N antennas, such that $N > N_{RF}$. A survey in [7] discussed various hybrid precoding structure in indoor as well as outdoor mmWave communications. As per IEEE 802.15.3c and IEEE 802.11ad standards, directional communication in mmWaves employs a two-level hierarchical search: a coarse broad beams (codebook design) search at RF level, followed by a fine directed beam search at baseband level. Although having a sub-array instead of a single antenna per RF unit in hybrid beamforming unit contributes to a higher beam directivity compared to digital beamforming, the associated joint RF-baseband signal processing is a computationally

This work was supported by the DST under Grant INT/Korea/P-46 and the DoT under Grant 4-23/5G-test-bed/2017-NT.

The authors are with the Department of Electrical Engineering and Bharti School of Telecommunication, Indian Institute of Technology Delhi, New Delhi 110016, India (e-mail: nancy.varshney@dbst.iitd.ac.in, swadesd@ee.iitd.ac.in).

complex problem [8]. Also, the inter-beam interference and power consumption at the BS are expected to increase with the increasing number of beams in a hybrid precoder [9].

The works on hybrid beamforming focus on scheduling UEs by providing a dedicated beam per UE, based on the assumption of sparse UE distribution, [7], [10]–[12]. Moreover, most of the prior works on mmWaves consider narrowband channels [8], [10], whereas practical mmWave channel is frequency-selective due to large bandwidth and beam squint effect. The few works on frequency-selective mmWave channel [11], [13]–[15] consider only one RF chain per UE, i.e., suggest allocating all the OFDM subcarriers to the same UE, which is not advantageous. In [13], different UEs are multiplexed over different subcarriers of the OFDM symbol, using fully-connected hybrid precoding. However, each UE baseband data stream is modulated using a separate RF chain, hence requiring $N_{RF} = M$, where M is the UE population. With the growing number of data-intensive devices, there can be thousands of mobile devices per square kilometer area where $M \gg N_{RF}$. A low-complexity selection combining (LCSC) algorithm was proposed in [14] to cater to high UE density. LCSC uses space-time multiplexing where it designates a subset of antenna elements of a single RF unit to each UE located within the beam's coverage area and multiplexes them over different time slots. This scheme suffers from the limitation of reduced beam directivity and coverage since at a particular time only a subset of total array elements is active.

In [15], the authors suggested grouping of UEs based on the correlation in angular dimension using joint spatial division multiplexing (JSDM) algorithm, with data of each UE being modulated by a separate RF chain. In this algorithm, the baseband-RF precoding of hybrid precoder having each of the N_{RF} RF chains connected to all the N antenna elements (fully-connected structure) is implemented in two-stages. As stated in [15], the selection of number of eigenmodes during pre-beamforming in JSDM is a non-trivial optimization problem. This approach gets complex with increased UE population, as it creates additional inter-group interference (a result of imperfect block diagonalization). The complexity of grouping M UEs in pre-beamforming algorithm at RF level increases with M as $\mathcal{O}(M^2)$. Besides high complexity, selection of appropriate N_{RF} is itself challenging while keeping hardware cost and power within budget.

We argue that there is a need to model a system of lower complexity that can provide reliable and fair services while optimally utilizing the available resources in mmWave multi-user cellular communications. As mentioned earlier, analog beamforming offers the most economical and reduced complexity designs; coupled with scheduling capabilities, it can provide satisfactory Quality-of-Service (QoS) to the UEs with high energy efficiency at the BS. In the context of analog beamforming, the authors in [3] surveyed several codebook designs and optimization techniques to find optimum beamwidth with reduced protocol overhead in a two-level exhaustive beam search procedure and also to reduce transmitter-receiver beam misalignment. Here also, these optimization problems target only indoor communication where single UE is served per RF chain, which is feasible only for small indoor population [16]. The performance of such schemes degrades in outdoor mmWave communication with high user density.

Another notable point is that most of the mmWave research focus on improving data rates and overcoming blockage effects through enhanced SNR (or radiated power density) by increasing transmit beamforming gain. For example, the approach in [17] estimates the optimal beamwidth that provides increased signal strength at the UE to overcome blockage effects. But increased power density causes elevated radiation hazard in the environment. The Federal Communications Commission (FCC) guidelines have defined limits on peak equivalent isotropically radiated power (EIRP) for communication over mmWaves to ensure health safety [18]. Accordingly, we are motivated to study the mmWave system performance while maintaining the same radiation level in the cell.

In this paper, we consider a 2-dimensional (2-D) scenario where a single analog beam serves a cell in a time-multiplexed fashion while keeping the EIRP level same in the cell, invariant of the antenna gain. The cell is sectorized into S similar disjoint sectors based on half-power beamwidth (HPBW) θ of the analog beam. S decreases with increasing θ, and vice versa. We aim to find S^* instead of θ^* because of mathematical convenience. The beam sweeps the area in either clockwise or counter-clockwise direction. As θ decreases, the total transmit power required to cover the same cell range per sector decreases, and the channel spatial reuse factor of the cell increases, whereas the time delay between the next schedule of each sector increases. Unlike in [15], the complexity of UE grouping algorithm in this model is independent of UE population. In the proposed scheme, multiple UEs are served simultaneously using OFDM in a single time slot per RF chain. Depending on UE density and distribution, a smaller value of θ may result in certain sectors having no UEs at all. In that case, the total cell sweep time can be reduced by not scheduling those sectors, thereby saving transmit energy. We ask the following question: What should be the optimum width of transmitted signal beam to cater to a specific UE distribution while jointly maximizing the average long-run UE data rate and energy efficiency of the BS?

To the best of our knowledge, this mmWave system behaviour has not been studied in the literature. Moreover, with huge bandwidth availability in mmWaves, employing time division multiple access scheduling to angularly spreaded UEs with large fixed time slots (e.g., 1 ms) may not be efficient [19]. We note that, in FCC guidelines, 5G systems can have variable-duration transmission units according to services and applications context. Thus, it is also worth studying the effect of variable scheduling based on different factors, such as UE distribution and channel conditions, on system performance. In [5] it was noted that, for static UEs with transmit beamforming, the position estimation error is inversely proportional to antenna array size. Along this line, with analog beamforming at BS, keeping EIRP constant irrespective of beamwidth, the effect of localization might have some role in deciding the optimal sector beamwidth in downlink mmWave communications. This is also an aspect of investigation in this work.

B. Contribution

Key contributions of this work are as follows:

a) This paper proposes and studies a prototype employing a single analog beam to cater to the outdoor UE population in wideband mmWave communications using a sectored-cell with time multiplexing approach. A narrow beam of optimally-chosen beamwidth, generated from an RF chain, serves multiple UEs during the beam sojourn time using the OFDMA scheme, as opposed to single UE per beam in the existing literature. The effect of channel frequency selectivity along with the minimum QoS requirements of each UE is considered in performance evaluation. As an extension, a comparison between a single beam system and multiple beams system is also studied; it is noted that while transmission over multiple analog beams may achieve a higher average long-run UE rate with respect to the single analog beam system, it is always at the cost of reduced BS energy efficiency.

b) An optimization problem is formulated to find the optimum sectorization scheme for the single analog beam system that maximizes the average long-run UE rate (average of all UEs' rates with each UE rate averaged over cell sweep time) and BS energy efficiency. It is observed that a Pareto optimal trade-off exists between average long-run UE rate and BS energy efficiency.

c) A method of reduced complexity is proposed to estimate optimal sectorization scheme for average long-run UE rate maximization for a given UE population.

d) For catering to non-homogeneous UE distribution using single beam, a variable-time scheduling per sector is formulated. The studies reveal that, for non-uniformly distributed UEs, variable time scheduling scheme provides considerably improved performance in terms of maximum achievable average long-run UE rate as well as BS energy efficiency.

e) Lower bounds on UE position estimation error is derived for analog beamforming using a practical antenna radiation pattern of the array. The impact of position error on the optimal beamwidth selection is also investigated.

C. Organization

System model is presented in Section II. Sector sum rate maximization and optimum sector beamwidth optimization problem along with analysis are presented in Section III. Simulation results are presented in Section IV, and concluding remarks are drawn in Section V.

Notations: For a vector $\mathbf{A}$, $A(n)$ indicates the n^{th} element of the vector. $|\cdot|$ and $\bigcup$ denote cardinality and set union, respectively. $\mathcal{O}$ denotes algorithm run-time complexity.

II. System Model

This section outlines the system model in a 2-D outdoor cellular scenario. Let there be a BS with one RF unit and M UEs $\{\mathrm{x}_1, \mathrm{x}_2, ..., \mathrm{x}_M\}$, distributed over a cell. Each UE x_i has a distribution according to some probability function $\mathcal{Q}(\cdot)$ in $\mathbb{R}^2$. Let $\mathcal{S}$ be a set; its elements denote the total number of sectors S in a cell when it is sectorized into equal and identical regions of angular width θ, i.e, $\mathcal{S} = \{S | S = \lceil 2\pi/\theta \rceil, \theta_{min} \leq \theta \leq \theta_{max}\}$. Let s denote the index of a particular sector, i.e., $s = \{1, \ldots, S\}$ when the total number of sectors is $S \in \mathcal{S}$. Then, the probability of finding a UE in the s^{th} sector of a given sectorization $S \in \mathcal{S}$ scheme given by $\mathcal{Q}(s)$ [20]. In case of uniform UE distribution, the probability of finding a UE in a sector is $\mathcal{Q}(s) = 1/S$ for $s = 1, \ldots, S$. To model non-uniform UE distribution a cell is divided into Z identically wide zones with the probability of distribution of each UE in the z^{th} zone being p_z, such that $\sum_{z=1}^{Z} p_z = 1$.

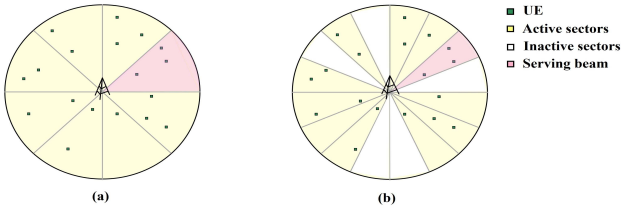

Figure 1. Pictorial representation of a 2-D single cell with 20 independent and uniformly distributed UEs, served by a single analog beam when the cell sectorization scheme is (a) $S = 8$ sectors, and (b) $S = 16$ sectors.

The UEs are considered to be equipped with a single omnidirectional antenna and BS has a uniform linear array (ULA) consisting of N identical isotropic antenna elements connected to an RF chain. Maximum gain of ULA is at steering angle $\phi = 0^o$ and is equal to NG_0, where G_0 is single antenna element gain. The beam gain G is assumed constant and equal to maximum gain within the sector and zero outside [16]. The HPBW of a beam is $\theta \approx 2/N$ radians, that divides the cell into $S = \lceil 2\pi/\theta \rceil$ similar disjoint sectors. Let P_{tot} be the total transmit power available at the BS. EIRP $(= P_{tot}G_0)$ of the BS is kept constant so that the cell coverage is static irrespective of the beam directivity. Therefore, to maintain the same SNR at UE at a distance d, the transmit power requirement in a sector of θ HPBW is calculated as:

$$P_t = \text{EIRP}/G = P_{tot}/N. \tag{1}$$

The channel model between UE and BS at mmWave is quite different from conventional cellular networks. There exists either a line-of-sight (LOS) link or non-line-of-sight (NLOS) link due to presence of large obstacles. Considering the study in [21], the probability of a link being in LOS for the k^{th} UE at distance d from BS is $\Pr(d) = 1/(1 + e^{d_2(d-d_1)})$, where $d_1 = 97$ and $d_2 = 0.0034$. Let $\gamma_{n,k}(d)$ be a parameter that denotes the combined channel state and location information of the k^{th} UE, at a distance d, over the n^{th} subcarrier, which is expressed as:

$$\gamma_{n,k}(d) = \frac{|h^{n,k}_{LOS}|^2 PL^{n,k}_{LOS} \Pr(d) + |h^{n,k}_{NLOS}|^2 (1 - \Pr(d)) PL^{n,k}_{NLOS}}{N_0 B / N_c G G_r}. \tag{2}$$

N_c is the total number of subcarriers from a set of subcarriers denoted by $\mathcal{N}_c$, G and G_r are respectively transmitter and receiver antenna gain, N_0 is the noise power spectral density, B is the total channel bandwidth, $h^{n,k}_{LOS}$ and $h^{n,k}_{NLOS}$ are small scale LOS and NLOS channel parameters of k^{th} UE over n^{th} subcarrier that are respectively Rician and Rayleigh distributed, and $PL^{n,k}_{LOS}$ and $PL^{n,k}_{NLOS}$ are the LOS and NLOS components of path loss of k^{th} UE at a distance d.

When a receiver receives signals from various directions,

narrow beam increases effective path loss, as some of the MPCs get blocked due to reduced beamwidth. This decreases the overall signal power reception. Thus, the beamwidth dependent path loss model, for both LOS and NLOS link, is $PL = PL(d) + \Delta PL(\theta)$ [dB]. Here, $PL(d)$ denotes the beamwidth independent path loss and $\Delta PL(\theta)$ denotes the additional path loss due to loss of MPCs with narrow beamwidth. The angle independent path loss for LOS and NLOS link are obtained as [22]:

$$\begin{aligned} PL_{LOS}(d) &= 61.4 + 20\log_{10}(d) + \mathcal{N}(0, 33.64) \text{ [dB]} \\ PL_{NLOS}(d) &= 72.0 + 2.92 \times 10\log_{10}(d) + \mathcal{N}(0, 75.69) \text{ [dB]}. \end{aligned} \tag{3}$$

$\Delta PL(\theta)$ at a fixed operating carrier frequency is $\Delta PL(\theta) = \zeta\,[\pi/(180^o \times \theta) - 1/360^o]$ [23], where ζ is a constant for LOS (ζ_{LOS}) and NLOS (ζ_{NLOS}).

Thus, the data rate $r_{n,k}$ for the k^{th} UE over the n^{th} subcarrier is $r_{n,k} = (B/N_c)\log_2(1+P_{n,k}\gamma_{n,k})$, where $P_{n,k}$ is the power of k^{th} UE over n^{th} subcarrier. In order to perform water-filling over the subcarriers, perfect channel knowledge of all UEs is considered at the BS.

III. Problem Formulation and Analysis

This section describes the problem formulation for optimal sectorization scheme S^* (or optimal sector beamwidth θ^*) in a cell under different scenarios. Firstly, we estimate S^* that maximizes the average long-run UE rate and BS energy efficiency, assuming perfect UE location knowledge at the BS. Next, when the BS has only an estimated position of the UEs using analog beamforming, we derive a lower bound on localization error and incorporate it to find S^*.

A. Optimal sectorization scheme with perfect UE location information

Here we present the steps involved in estimating S^* for the model described in Section II. For a sectorization scheme, $S \in \mathcal{S}$ each of the sectors has a full channel for its use. Hence, the sum rate of a sector during its schedule time, i.e., instantaneous sum rate, can be maximized independently. Consequently, the problem of finding optimal beamwidth can be divided into two parts. First, for a given value of S, we find the instantaneous sum rate for all of the active sectors (sectors with at least one UE) independently. After that, using an exhaustive search over all sectorization schemes $S \in \mathcal{S}$, we estimate S^* that maximizes the average long-run UE rate and BS energy efficiency. Further, we provide a reduced complexity method that estimates a sub-optimal sectorization scheme $\hat{S}^*$ for a given UE population without exhaustive search.

1) Instantaneous sum rate maximization per sector

Instantaneous sum rate maximization problem using OFDM in s^{th} sector of a sectorization scheme $S \in \mathcal{S}$, with minimum instantaneous rate constraint R_0 per UE (except for the UEs in outage) and maximum available transmit power P_t (given by (1)), is formulated as:

$$\begin{aligned} (\mathcal{P}1): \max_{\pi_{n,k}, P_{n,k}} \quad & \sum_{k=1}^{K_s}\sum_{n=1}^{N_c} r_{n,k}\pi_{n,k} \\ \text{s.t.} \quad & C11: \sum_{k=1}^{K_s} \pi_{n,k} \le 1, && \forall n \\ & C12: \sum_{n=1}^{N_c} r_{n,k}\pi_{n,k} \ge R_0, && \forall k \\ & C13: \pi_{n,k} \in (0,1), && \forall n,k \\ & C14: \sum_{k=1}^{K_s}\sum_{n=1}^{N_c} P_{n,k} \le P_t \\ & C15: P_{n,k} \ge 0, && \forall n,k \end{aligned} \tag{4}$$

where the UE$-$subcarrier association variable $\pi_{n,k} = 1$ if k^{th} UE assigned n^{th} subcarrier; 0 otherwise. The resources are allocated only to $K_s \subseteq K_s'$ UEs that are not in complete outage. Here, K_s' is the set of UEs falling in the s^{th} sector such that $\bigcup_{s\in S} K_s' = M$. We define k^{th} UE to be in outage if $\max_{n\in\mathcal{N}_c}\{\gamma_{n,k}\} < \gamma_{th}$. Here, γ_{th} is the threshold channel state value, $k = 1, 2, \cdots, K_s$.

$\mathcal{P}1$ jointly allocates subcarriers and power to maximize the achievable sum rate and is a mixed-integer non-convex programming problem. Such problems are Non-deterministic Polynomial time (NP)-hard and not computationally efficient. The Lagrangian of $\mathcal{P}1$ is formulated as:

$$\begin{aligned} \mathcal{L}(\pi_{n,k}, P_{n,k}, \bar{\eta}_k, \bar{\lambda}) &= \sum_{k=1}^{K_s}\sum_{n=1}^{N_c} r_{n,k}\pi_{n,k} + \\ &\sum_{k=1}^{K_s} \bar{\eta}_k \left(\sum_{n=1}^{N_c} r_{n,k}\pi_{n,k} - R_0\right) - \bar{\lambda}\left(\sum_{k=1}^{K_s}\sum_{n=1}^{N_c} P_{n,k} - P_t\right) \\ &= \sum_{n=1}^{N_c}\sum_{k=1}^{K_s}\left(r_{n,k}\pi_{n,k}\left(1+\bar{\eta}_k\right) - \bar{\lambda}P_{n,k}\right) - \sum_{k=1}^{K_s}\bar{\eta}_k R_0 + \bar{\lambda}P_t \\ &= \sum_{n=1}^{N_c}\mathcal{L}_n\left(\pi_{n,k}, P_{n,k}, \bar{\eta}_k, \bar{\lambda}\right) - \sum_{k=1}^{K_s}\bar{\eta}_k R_0 + \bar{\lambda}P_t \end{aligned} \tag{5}$$

where $\bar{\lambda}$ is the Lagrange multiplier for power constraint $C14$ and $\bar{\eta}_k\ \forall k$ are the Lagrange multipliers for rate constraint $C12$. In order not to violate complementary slackness condition $\bar{\lambda} \ge 0, \bar{\eta}_k \ge 0\ \forall k$. The binary variable constraints $C11, C13$, and positive power constraint $C15$ are not included into the Lagrangian formulation to circumvent unnecessary complexity but are considered at later stage. This method will not alter the solution. Further, the optimization problem satisfies the time-sharing condition [24]. Hence, the duality gap is zero, i.e., the dual function solution is same as the primal function solution. The Lagrangian dual function is:

$$\begin{aligned} g(\bar{\eta}_k, \bar{\lambda}) = \max_{\pi_{n,k}, P_{n,k}} \quad & \mathcal{L}(\pi_{n,k}, P_{n,k}, \bar{\eta}_k, \bar{\lambda}) \\ \text{s.t.:} \quad & C11, C13, C15. \end{aligned} \tag{6}$$

The above equation can be decomposed into the following N_c

independent problems as:

$$g'(\bar{\eta}_k, \bar{\lambda}) = \max_{P_{n,k}} \quad \mathcal{L}_n(\pi_{n,k}, P_{n,k}, \bar{\eta}_k, \bar{\lambda}) \\ \text{s.t.:} \quad C15. \tag{7}$$

Assuming k^{th} UE is active on n^{th} subcarrier, (7) reduces to

$$g'(\bar{\eta}_k, \bar{\lambda}) = \max_{P_{n,k}} \quad \pi_{n,k}\frac{B(1+\bar{\eta}_k)}{N_c}\log_2(1+P_{n,k}\gamma_{n,k}) - \bar{\lambda}P_{n,k} \\ \text{s.t.:} \quad C15. \tag{8}$$

(8) is concave in $P_{n,k}$. Hence by Karush-Kuhn-Tucker (KKT) condition the solution to optimal power allocation $P'_{n,k}$ is:

$$\frac{\partial \mathcal{L}_n(\pi_{n,k}, P_{n,k}, \bar{\eta}_k, \bar{\lambda})}{\partial P_{n,k}} = \frac{(B/N_c)\pi_{n,k}\gamma_{n,k}(1+\bar{\eta}_k)}{\ln 2(1+P_{n,k}\gamma_{n,k})} - \bar{\lambda} = 0. \tag{9}$$

Incorporating the constraint $C15$, from (9) we have

$$P'_{n,k} = \left[\frac{(B/N_c)\pi_{n,k}(1+\bar{\eta}^*_k)}{\ln 2\bar{\lambda}^*} - \frac{1}{\gamma_{n,k}}\right]^+ \tag{10}$$

where $[x]^+ = \max(x, 0)$, $\bar{\lambda}^*$ is the optimal dual variable corresponding to power constraint, and $\bar{\eta}^*_k$ is the optimal dual variable corresponding to minimum rate constraint of the k^{th} UE.

The Lagrangian variables are updated using sub-gradient based approach [25] as follows:

$$\bar{\lambda}^{(i+1)} = \left[\bar{\lambda}^{(i)} - \bar{\epsilon}^{(i)}\left(P_t - \sum_k\sum_n P^*_{n,k}\right)\right]^+ \\ \bar{\eta}^{(i+1)}_k = \left[\bar{\eta}^{(i)}_k - \bar{\epsilon}^{(i)}\left(\sum_n \pi^*_{n,k} r_{n,k} - R_0\right)\right]^+ \forall k \tag{11}$$

where $\bar{\epsilon}^{(i)}$ is the step size at i^{th} iteration and is chosen according to the constant step size policy, independent of iteration number. Update of $\bar{\eta}_k$ ensures the minimum rate constraint. Finally, the UE−subcarrier combination is selected that offers the maximum sum rate while satisfying minimum rate constraint of all the UEs.

Proposed sub-optimal algorithm: The above exhaustive search method requires the execution of water-filling algorithm for at least $K_s^{N_c}$ times. To address this, we convert the optimization problem $\mathcal{P}1$ into a sub-optimal problem of lower complexity, leveraging the fact that sum rate maximization incurs a negligible penalty when only channel adaptive subcarrier allocation is performed [26]. The steps involved in sub-optimum optimization approach are:

(i) To reduce complexity at a starting resource allocation step, we divide the total transmit power P_t equally among all the subcarriers, i.e., if $\pi_{n,k} = 1$ then $P_{n,k} = P_t/N_c$.

(ii) Initially we allocate each subcarrier to a UE with the best channel condition without considering the rate constraints $C12$ [26]. Hence, the initial optimal UE–subcarrier association to the n^{th} subcarrier is found as $k'(n) = \underset{k}{\text{argmax}}\ r_{n,k}$.

(iii) Subcarrier allocation from step (ii) does not satisfy the constraint $C12$. Therefore, we carry out the subcarrier allocation on a user-by-user basis for all UEs to satisfy the minimum rate constraints (unless they are in complete outage) using cost function approach [26]:

$$c_{n,k} = \frac{(r_{n,k'(n)} - r_{n,k})}{r_{n,k}}$$

where $k'(n)$ is the user that is initially assigned to n^{th} subcarrier from step (ii). The cost function is proportional to the reduction in overall sum rate. For k^{th} UE that does not satisfy the constraint $C12$, the $\tilde{n}^{th}$ subcarrier assigned to meet its minimum rate requirements is found as $\tilde{n} = \underset{n}{\text{argmin}}\ c_{n,k}$. Further, during subcarrier reallocation the $\tilde{n}^{th}$ subcarrier that was initially associated to $k'(\tilde{n})^{th}$ UE should not be reallocated to another UE if this reallocation violates the minimum rate constraint of the original UE, i.e., $\tilde{n}^{th}$ subcarrier is not allocated to k^{th} UE if $R(k') - r_{\tilde{n},k'} < R_0$, where $R(k') = \sum_{n=1}^{N_c}\pi_{n,k'}r_{n,k'}$.

Remark 1. *$\mathcal{P}1$ has a feasible solution if $K_s \leq N_c$ and the value of R_0 is appropriately chosen. In case $K_s > N_c$ the subcarriers are assigned to only N_c UEs having best channel conditions among K_s UEs, with minimum rate requirement fulfilment of only N_c UEs while the remaining $K_s - N_c$ UEs will be considered in outage in that cycle of sector sweep.*

Remark 2. *When the value of K_s is very low, the subcarrier allocation algorithm does not utilize the channel capacity to its maximum, as there may be some subcarriers that are not assigned to any UE due to deep fade. When the K_s increases, the probability of a subcarrier being in deep fade simultaneously for all the UEs decreases, thereby increasing overall sum rate (channel utilization). However, when the value of K_s is sufficiently high, assigning a subcarrier to a UE whose rate constraint was not satisfied earlier (from step (ii)), instead of the UE with best channel gain over the subcarrier, leads to a decrease in overall sum rate.*

(iv) During UE–subcarrier assignment in steps (ii) and (iii), equal power allocation per subcarrier is considered, as given in step (i). Hence, in order not to violate minimum rate constraints of UEs while executing water-filling, each UE is offered a share of total transmit power proportional to the number subcarriers assigned to it. Therefore, the problem of power assignment to UEs further reduces to solving power allocation to each UE over its set of assigned subcarriers, independent of other UEs. Hence, the power assignment optimization of k^{th} UE over its assigned set of subcarriers Ω_k (step (iii)) is expressed as:

$$(\mathcal{P}2): \max_{P_{n,k}} \quad \sum_{n\in\Omega_k} r_{n,k} \\ \text{s.t.} \quad C21: \sum_{n\in\Omega_k} P_{n,k} \leq \frac{|\Omega_k|P_t}{N_c} \\ C22: P_{n,k} \geq 0, \forall n \in \Omega_k. \tag{12}$$

The Lagrangian of $\mathcal{P}2$ is formulated as:

$$\mathcal{L}_k(\{P_{n,k}\}_{\forall n\in\Omega_k},\lambda_k) = \sum_{n\in\Omega_k} r_{n,k} - \lambda_k \sum_{n\in\Omega_k}\left(P_{n,k} - \frac{|\Omega_k|P_t}{N_c}\right) \quad (13)$$

where λ_k is the dual variable associated with maximum power constraint $C21$.
Since (13) is concave in $P_{n,k}$, from KKT conditions it follows:

$$\frac{\partial\mathcal{L}_k(\{P_{n,k}\}_{\forall n\in\Omega_k},\lambda_k)}{\partial P_{n,k}} = \frac{(B/N_c)\pi_{n,k}\gamma_{n,k}}{\ln 2(1+P_{n,k}\gamma_{n,k})} - \lambda_k = 0 \quad (14)$$

$$P^*_{n,k} = \left[\frac{(B/N_c)\pi_{n,k}}{\ln 2\lambda^*_k} - \frac{1}{\gamma_{n,k}}\right]^+ \quad (15)$$

where λ^*_k is the optimal dual variable corresponding to power constraint of the k^{th} UE and is found using sub-gradient method with sufficiently small constant step size $\hat{\epsilon}$, given as:

$$\lambda^{(i+1)}_k = \left[\lambda^{(i)}_k - \hat{\epsilon}^{(i)}\left(\frac{|\Omega_k|P_t}{N} - \sum_k\sum_n P^*_{n,k}\right)\right]^+. \quad (16)$$

Steps (i) to (iv) solve the original constrained problem in a sub-optimum way. The optimal solution of $\mathcal{P}1$ is an upper bound on the sub-optimum solution obtained using $\mathcal{P}2$. Both the solutions show similar behaviour on varying parameters, which we have verified via simulation in Section IV. Thus, to calculate the instantaneous rate of all the UEs in the active sectors sub-optimum approach is used that significantly reduces computation complexity.

Remark 3. *In this paper, S^* is estimated by comparing system performance over all sectorization schemes $S \in \mathcal{S}$, which is a relative analysis. Hence, the qualitative nature of the solution of $\mathcal{P}1$ is of more interest than its quantitative nature. Therefore, the instantaneous rate allocation can be done with other resource allocation techniques as well.*

2) Sector beamwidth optimization

We use the average long-run UE rate to evaluate the QoS of the system from UE's perspective. Let t_s be the time spent by the beam in s^{th} sector and $R_{k,s}$ be the instantaneous rate of k^{th} UE in the s^{th} sector during beam sojourn time t_s of the sectorization scheme $S \in \mathcal{S}$. Also, let $\mathcal{S}_I(S) = \{s|K_s \neq 0, \forall s \in \{1,\ldots,S\}\}$ denote the support of s of a particular $S \in \mathcal{S}$. Then long-run data rate $\hat{R}_{k,s}$ of k^{th} UE in s^{th} sector is:

$$\hat{R}_{k,s} = \frac{R_{k,s}t_s}{\sum_{s\in\mathcal{S}_I(S)} t_s}. \quad (17)$$

Hence, the average of long-run UE rate in sectorization scheme S is:

$$\bar{R}(S) = \frac{1}{M}\sum_{s\in\mathcal{S}_I(S)}\sum_{k=1}^{K_s}\hat{R}_{k,s}. \quad (18)$$

Solving for the optimal resource allocation for sum rate maximization with rate constraints result in non-uniform rate distribution among the UEs. Therefore, we use Jain's fairness index [27] to characterize fairness of the long-run UE rate distribution. In general it is defined as:

$$J(S) = \frac{1}{M}\frac{(\sum_{j=1}^{j=M}\hat{R}_j)^2}{\sum_{j=1}^{j=M}(\hat{R}_j)^2} \quad (19)$$

where $\hat{R}_j$ is the long-run UE rate of j^{th} UE. Another parameter of interest is the energy efficiency at BS achieved by a particular sectorization scheme $S \in \mathcal{S}$. Energy efficiency is defined as the ratio of total number of bits transmitted to the total energy consumed in transmitting those bits. For a sectorization scheme S, it is given as:

$$\bar{E}(S) = \frac{\sum_{s\in\mathcal{S}_I(S)}\sum_{k=1}^{K_s}(R_{k,s}t_s)}{\sum_{s\in\mathcal{S}_I(S)}(\sum_{n=1}^{N_c}\sum_{k=1}^{K_s}P_{n,k}t_s)}. \quad (20)$$

To find S^* that maximizes $\bar{R}$ and $\bar{E}$, the multi-objective optimization (MOP) is expressed as:

$$(\mathcal{P}3)\quad \text{MOP}: \quad \underset{S\in\mathcal{S}}{\text{argmax}}[\bar{R}(S)\ \ \bar{E}(S)]. \quad (21)$$

Both average long-run UE rate $\bar{R}$ and BS energy efficiency $\bar{E}$ are functions of S. It is possible that the two objectives are of conflicting nature, i.e., in choosing the optimal sectorization scheme S^*, there exists a trade-off between $\bar{E}$ and $\bar{R}$. In such a situation the solution to MOP problem consists of a complete set of dominated solutions, known as the Pareto Set.

Definition: A decision vector $S^* \in \mathcal{S}$ is Pareto-optimal if there exists no other feasible decision vector $\tilde{S} \in \mathcal{S}$ that dominates S^*, i.e., S^* is strictly better than $\tilde{S}$ in all objectives.

In order to find Pareto-optimal points, we use single objective optimization problem (SOP) approach [28] for a given $\bar{R}$ (i.e., fixing the value of one of the objectives) as shown below:

$$(\mathcal{P}4)\quad \text{SOP}(\bar{R}): \underset{S\in\mathcal{S}}{\text{argmax}}\ \bar{E}(S). \quad (22)$$

For a given value of $\bar{R}$, let S^* be an optimal solution of $\mathcal{P}4$, with corresponding objective value $\bar{E}^*$. Then S^* is a Pareto-optimal solution to MOP. This process is repeated for different values of $\bar{R}$, and thus a set of Pareto optimal $(\bar{R},\bar{E}^*)$ combinations is achieved.

Proposed reduced-complexity method: The method discussed so far performs exhaustive search over all $S \in \mathcal{S}$ in order to determine the optimal number of sectors S^* that maximizes average long-run UE rate for a given UE population. This procedure needs to be re-executed if the UE population changes. To reduce online computation requirements, we propose a reduced-complexity method for estimating sub-optimal $\hat{S}^*$ that achieves maximum average long-run UE rate without having to solve the complete optimization problem when the UE population changes.

In fixed time scheduling, let $t_1 = t_2 = \cdots = t_S = T$, then total cell sweep time is $\sum_{s\in\mathcal{S}_I(S)} t_s = T \times |\mathcal{S}_I(S)|$. From (17) and (18), average long-run UE rate is:

$$\bar{R}(S) = \frac{\sum_{s\in\mathcal{S}_I(S)}\sum_{k=1}^{K_s}R_{k,s}t_s}{M\sum_{s\in\mathcal{S}_I(S)}t_s} = \frac{\sum_{s\in\mathcal{S}_I(S)}\sum_{k=1}^{K_s}R_{k,s}t_s}{M\times T\times|\mathcal{S}_I(S)|}. \quad (23)$$

Since it is assumed that a given value of $S \in \mathcal{S}$ partitions the cell into identical disjoint sectors, the resource allocation of a sector is independent of other. Another notable point is that the individual sector's sum rate is a function of the number of UEs K_s present in s^{th} sector (see *Remark* 2) and of sectorization scheme S (because of variation of path loss with θ).

Let for a sectorization scheme $S \in \mathcal{S}$ and M uniformly distributed UEs, each active sector has $K_s \approx M/|\mathcal{S}_I(S)|$ UEs with corresponding sector sum rate R_{sum} during T. Then, (23) can be approximated as:

$$\bar{R}(S) \approx \frac{|\mathcal{S}_I(S)| \times R_{sum} \times T}{M \times T \times |\mathcal{S}_I(S)|} = \frac{R_{sum}}{M}. \tag{24}$$

Hence, maximization of the average long-run UE rate is essentially to maximize the sum rate of each sector for a given UE population. To maximize R_{sum} we need to solve the optimization problem $\mathcal{P}1$ for $1 \leq K_s \leq N_c$. But $\mathcal{P}1$ is an NP-hard problem. Therefore, we solve it in a sub-optimum way as described in Section III-A1. Thereafter, to find sub-optimal sectorization scheme $\hat{S}^*$ with reduced complexity method, we use a dictionary based approach. In this, a *dictionary* containing the sector sum rates corresponding to all combinations of sectorization schemes and the number of UEs in a sector is pre-calculated at BS using the statistical channel and UE distribution. The dictionary is then used to directly find $\hat{S}^*$ for any given UE population.

The maximum number of UEs supported per sector is equal to the number of subcarriers. Let the dictionary $\mathcal{R}_{\mathbf{sum}}$ be a matrix of sector sum rates with dimensions equal to $|\mathcal{S}| \times N_c$. Hence, $\mathcal{R}_{\mathbf{sum}}(S, K_s)$ corresponds to the sum rate when the sectorization scheme is S and number of UEs in a sector is K_s. Once $\mathcal{R}_{\mathbf{sum}}$ is known, $\hat{S}^*$ for M uniformly distributed UEs is found as:

$$\hat{S}^* = \underset{S \in \mathcal{S}}{\operatorname{argmax}}\ \mathcal{R}_{\mathbf{sum}}\left(S, \left\lceil \frac{M}{S} \right\rceil\right). \tag{25}$$

Similarly, maximization of average long-run UE rate with non-uniformly distributed UEs is synonymous to maximizing throughput of the highest populated zone which is the driving factor of the overall cell throughput. So, in case of M non-uniformly distributed UEs $\hat{S}^*$ is found as:

$$\hat{S}^* = \underset{S \in \mathcal{S}}{\operatorname{argmax}}\ \mathcal{R}_{\mathbf{sum}}\left(S, \left\lceil \frac{Mp^{max}}{S/Z} \right\rceil\right) \tag{26}$$

where Z is the total number of zones in a cell, and $p^{max} = \max_z\{p_z\}$ where $z = 1, 2 \ldots, Z$.

3) Variable time scheduling

To account for the possibility of non-uniform UE distribution in a cell we study two scheduling schemes, namely, fixed time scheduling and variable time scheduling. In fixed time scheduling, $t_s = T$ if $K_s \neq 0$; 0 otherwise. In variable time scheduling, the schedule time t_s of each active sector is a function of the number of UEs in the that sector, i.e., $t_s \propto {K_s}^{\alpha}$, where α is the shaping parameter that governs the nature of scheduling time. $\alpha = 0$ corresponds to a fixed time scheduling scheme. Besides, we observe that as θ decreases the number of UEs located within a sector decreases. Also, the spatial frequency reuse factor of the cell is improved. As a result, UEs benefit from spatial multiplexing gain which provides improved fairness among UEs.

Lemma 1: For a fixed α and $K_s \leq N_c$, Jain's fairness index J is an increasing function of S, when resource allocation is performed using sum rate maximization approach.

Proof: Given in the Appendix A.

We consider fairness index of fixed time scheduling scheme $J_{\alpha_0}(S^*)$ at $\alpha = \alpha_0 = 0$ as a benchmark for the variable time scheduling. Here, S^* is the optimal sectorization scheme such that $S^* = \underset{S \in \mathcal{S}}{\operatorname{argmax}}\ \bar{R}_{\alpha_0}(S)$. Thus, the benchmark of the maximum average long-run UE rate and energy efficiency are $\bar{R}_{\alpha_0}(S^*)$ and $\bar{E}_{\alpha_0}(S^*)$, respectively. Algorithm 1 and algorithm 2 find the optimum shaping parameter α_R and α_E that maximizes the average long-run UE rate and energy efficiency, respectively, without compromising on fairness achieved with fixed scheduling scheme at S^*. Let S'^* and S''^* be the optimal sectorization scheme at α_R obtained using algorithm 1 and α_E obtained using algorithm 2, respectively.

The input to the Algorithm 1 is the instantaneous rate $R_{k,s}$ of all UEs, obtained in Section III-A1; $\bar{R}_\alpha(S)$, $J_\alpha(S)$, and $\bar{E}_\alpha(S)$ are calculated using the relation $t_s \propto K_s^\alpha$ in (18), (19), and (20), respectively for all possible sectorization schemes $S \in \mathcal{S}$. Next, a counter j is initialized with a suitable step size $\hat{\delta}$. For the j^{th} iteration, we search for α_R (if any) such that $\bar{R}_{\alpha_R}(S'^*) > \bar{R}_{\alpha_0}(S^*)$ while maintaining the required conditions $J_{\alpha_R}(S'^*) \geq J_{\alpha_0}(S^*)$ and $\bar{E}_{\alpha_R}(S'^*) \geq \bar{E}_{\alpha_0}(S^*)$. The iteration continues to increment α_j and update α_R till a stopping condition is met (as given in the algorithms). The stopping condition is explained as follows. When α is sufficiently high, variable time scheduling scheme allocates more time to the sectors with high value K_s and nearly zero time to sparsely populated sectors, thus degrading the fairness, i.e., $J_\alpha(S) < J_{\alpha_0}(S)\ \forall S$. However, from Lemma 1 we know that for a fixed value of α the fairness index $J_\alpha(S) \geq J_\alpha(S^*)$ for $S > S^*$. Combining the above two deductions, the feasible search region of α_R, given a benchmark fairness index, is given by the condition $J_\alpha(S = 360) \geq J_\alpha(S'^*) \geq J_{\alpha_0}(S^*)$

Algorithm 1: α_R search algorithm

```
Input: R_{k,s}  ∀S ∈ S
Output: α_R
Initialize: j = 1, α_R = α_0 = 0, δ̂ = 0.1;
Find: S* = argmax_{S∈S} R̄_{α_0}(S), R_opt ← R̄_{α_0}(S*);
while true do
    α_j = α_{j−1} + δ̂;
    Compute: R̄_{α_j}(S), J_{α_j}(S), Ē_{α_j}(S)  ∀S;
    if J_{α_j}(S = 360) < J_{α_0}(S*) then
        break; (Stopping condition)
    end
    for i = S* : 360 do
        if J_{α_j}(i) ≥ J_{α_0}(S*) & R̄_{α_j}(i) > R_opt &
          Ē_{α_j}(i) ≥ Ē_{α_0}(S*) then
            α_R = α_j, R_opt = R̄_{α_j}(i), S'* = i;
        end
    end
    increment j;
end
```

Algorithm 2: α_E search algorithm

Input: $R_{k,s}\ \ \forall S \in \mathcal{S}$
Output: α_E
Initialize: $j = 1, \alpha_E = \alpha_0 = 0,\ \hat{\delta} = 0.1$;
Find: $S^* = \underset{S\in\mathcal{S}}{\text{argmax}}\ \bar{R}_{\alpha_0}(S), \bar{E}_{opt} = \bar{E}_{\alpha_0}(S^*)$;
while *true* **do**
  $\alpha_j = \alpha_{j-1} + \hat{\delta}$;
  Compute: $\bar{E}_{\alpha_j}(S),\ J_{\alpha_j}(S), \bar{R}_{\alpha_j}(S)\ \ \forall S$;
  if $J_{\alpha_j}(S = 360) < J_{\alpha_0}(S^*)$ **then**
    break; (*Stopping condition*)
  end
  for $i = S^* : 360$ **do**
    if $J_{\alpha_j}(i) \geq J_{\alpha_0}(S^*)$ & $\bar{E}_{\alpha_j}(i) > \bar{E}_{opt}$ & $\bar{R}_{\alpha_j}(i) \geq \bar{R}_{\alpha_0}(S^*)$ **then**
      $\alpha_E = \alpha_j, \bar{E}_{opt} = \bar{E}_{\alpha_j}(i), S''^* = i$;
    end
  end
  increment j;
end

(we take the maximum possible number of sectors $S = 360$). Therefore, during the search for α_R in Algorithm 1 we keep incrementing α_j till we have $J_{\alpha_j}(S = 360) \geq J_{\alpha_0}(S^*)$; otherwise the optimal shape parameter search algorithm stops. Similarly, Algorithm 2 describes the steps to find the shaping parameter α_E that maximizes BS energy efficiency without compromising on the average long-run UE rate and the fairness achieved using fixed scheduling at S^*.

Remark 4. *In uniform UE distribution, difference in the values of K_s is minimal, i.e., its variance is nearly zero. Thus, the scheduled time of each sector is approximately the same and is equivalent to the fixed time scheduling scheme. In contrast, the variance of K_s is significant for non-uniformly distributed UEs within the cell.*

B. Optimal sectorization scheme with estimated UE position information

In the above study, we assumed that gain of the beam within a sector is NG_0 and zero outside. In practice, the array factor vector of a broadside ULA with half wavelength spacing is [29]:

$$\mathbf{a}(\hat{\theta}) = \frac{1}{\sqrt{N}}[1\ \ e^{j(k\bar{d}\sin(\hat{\theta}))} \dots e^{j(N-1)(k\bar{d}\sin(\hat{\theta}))}]^T, \quad \text{such that } \|\mathbf{a}(\hat{\theta})\|_2 = 1 \tag{27}$$

where $\bar{d}$ is the spacing between antenna elements, $\hat{\theta}$ is the angle deviation from the direction of maximum array gain, and $k\bar{d} = \pi$ is the uniform phase shift between two antenna elements. Thus, the resulting gain pattern of the ULA with the array factor given by (27) at an angle $\hat{\theta}$ is:

$$G(\hat{\theta}) = N\left|\left[\frac{1}{N}\frac{\sin(\frac{N\pi}{2}\sin(\hat{\theta}))}{\sin(\frac{\pi}{2}\sin(\hat{\theta}))}\right]\right|^2, \text{ for } -\pi \leq \hat{\theta} \leq \pi \tag{28}$$

with first null beamwidth (FNBW) of the broadside ULA being $\Theta = 2\left[\frac{\pi}{2} - cos^{-1}(2/N)\right]$. The derivation of ULA gain is provided in Appendix B.

The position estimation is assumed to be performed at UEs that are continuously listening to the channel for pilot symbols spread over the entire OFDM symbol bandwidth. This data is then reported back to the BS. Also, we assume perfect time synchronization between the transmitter and receiver. We define effective AOA (EAOA), denoted by Ψ, as the effective angle of signal arrival at the UE with respect to the beam steering direction of the serving sector. As a result, $\gamma_{n,k}$ given by (2) is will be a function of EAOA as well, in addition to distance d from BS, i.e.,

$$\gamma_{n,k}(d,\Psi) = \frac{|h^{n,k}_{LOS}|^2 PL^{n,k}_{LOS}\Pr(d) + |h^{n,k}_{NLOS}|^2(1-\Pr(d))PL^{n,k}_{NLOS}}{N_0 B/N_c G(\Psi)G_r} \tag{29}$$

This parameter is used at BS for resource allocation purpose as described in Section III-A1. Due to error in angle estimation, the estimated EAOA of UE at BS is Ψ_e, with $|\Psi - \Psi_e| \geq 0$, as shown in Fig. 2. Therefore, $\gamma_{n,k}(d,\Psi_e)$ of k^{th} UE over n^{th} subcarrier is the effective channel and location parameter at BS, while $\gamma_{n,k}(d,\Psi)$ being the original value.

The UE position information variables are d and Ψ. Let τ be the time-of-flight of signal from BS to UE. In case of a noiseless ideal channel and $G_r = 1$, using TOA and DOA methods, (d,Ψ) are jointly estimated from the equations: $\tau = d/c$, and $P_r = P_t^{Pilot}G(\Psi)/(4\pi d/\lambda)^2$, where c is the speed of light, P_t^{Pilot} is the known transmitted pilot signal power, and P_r is the received power at distance d and angle Ψ. In case of a fading and noisy channel, there will be error in position estimation. At UE, we estimate the position-related parameters Ψ and τ from LOS received signal only, assuming no NLOS components are received. The channel gain β_m, for $m = 1, 2 \dots, N$, for a UE with Ψ EAOA and corresponding $\mathbf{a}(\Psi)$ unit norm array response (given by (27)) is nearly same for all its elements, i.e., $\beta_m \approx \beta$. Then the channel vector is:

$$\mathbf{h} = \sqrt{N}\beta e^{-j2\pi f_c \tau}\mathbf{a}^H(\Psi). \tag{30}$$

At the BS, OFDM signal x is multiplied by array factor precoding vector $\mathbf{a}(\phi)$, where ϕ is ULA steering direction relative to the direction of maximum antenna gain. Since the transmitted power of the OFDM symbol with N transmitting antennas is P_t, the received signal is:

$$y = \sqrt{N}\beta e^{-j2\pi f_c \tau}\mathbf{a}^H(\Psi)\mathbf{s} + \nu \tag{31}$$

where $\mathbf{s} = \mathbf{a}(\phi)x$ is the transmitted signal with $E[\mathbf{s}^H\mathbf{s}] = P_t$, $\beta = \mathbb{E}_h\left[h\sqrt{\frac{(d_0/d)^l}{(4\pi d_0/\lambda)^2}}\right]$, h is small scale fading parameter of

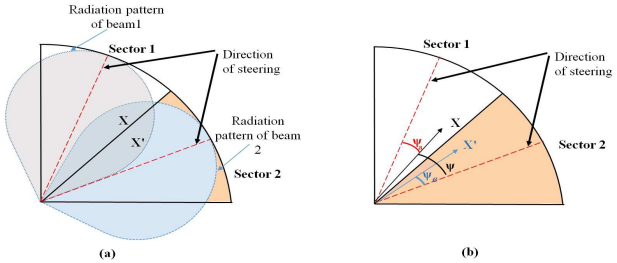

Figure 2. (a) Pictorial representation of true position $\mathbf{X}$ and estimated position $\mathbf{X}'$ of a UE along with the actual beam coverage region considering practical radiation pattern. (b) Demonstration of EAOA of the UE; Ψ is the true EAOA and Ψ_e is the estimated EAOA with respect to sector 2, whereas Ψ_a is the true EAOA of UE with respect to sector 1.

each subcarrier of OFDM symbol, d_0 is the reference distance, l is the LOS path loss exponent, and ν is the noise with distribution $\mathcal{N}(0, \sigma^2)$.

Here, other than τ and Ψ all other parameters are known apriori. Let $\Gamma = [\tau \quad \Psi]^T \in \mathcal{R}^{2\times 1}$ denote the vector of unknown channel parameters. We define $\hat{\Gamma} = [\tau_e \quad \Psi_e]^T$ to be an unbiased estimator of Γ, where τ_e and Ψ_e are unbiased estimator of TOA and EAOA (Fig. 2). Also, let $\mathcal{G}$ denote the number of transmissions required for UE position estimation using narrow beams. Then the lower bound on position error variance with unbiased estimator $\hat{\Gamma}$ is provided by the Cramer-Rao Lower Bound (CRLB) [30] as:

$$\text{var}(\hat{\Gamma} - \Gamma) \geq -\left[\mathbb{E}_{\mathbf{Y}|\Gamma}\left(\frac{\partial^2 \log f(\mathbf{Y}|\Gamma)}{\partial\Gamma\partial\Gamma^T}\right)\right]^{-1} \equiv (\mathbf{F}(\Gamma))^{-1} \tag{32}$$

where $\mathbf{Y} = [y^1, \ldots, y^{\mathcal{G}}]^T$, $\mathbb{E}_{\mathbf{Y}|\Gamma}[\cdot]$ denotes parametrized expectation function for Γ, $f(\mathbf{Y}|\Gamma)$ is the conditional likelihood of $\mathbf{Y}$ given Γ, and $\mathbf{F}(\Gamma)$ is the Fisher Information Matrix (FIM).

Lemma 2: The FIM for the received signal is given as:

$$\mathbf{F}(\Gamma) = \frac{1}{N_0 B}\sum_{g=1}^{\mathcal{G}} \text{Re}(\nabla y^g \nabla^H y^g). \tag{33}$$

Proof: Given in Appendix C.

Here $\text{Re}(\cdot)$ denotes real part operator, and ∇y^g is gradient of y^g with respect to Γ, given as:

$$\nabla y^g = \begin{bmatrix} -j2\pi f_c \sqrt{N}\beta \exp\{(-j2\pi f_c \tau)\}\mathbf{a}^H(\Psi)\mathbf{a}(\phi)x^g \\ \frac{-j2\pi\bar{d}}{\lambda} cos\Psi \sqrt{N}\beta \exp\{(-j2\pi f_c \tau)\}\mathbf{a}^H(\Psi)\mathbf{\Lambda}\mathbf{a}(\phi)x^g \end{bmatrix}. \tag{34}$$

$\mathbf{\Lambda} = \text{diag}[0 \ 1 \ldots (N-1)]$. We assume that the BS has a separate ULA for each sector and the beam in each sector is steered in the direction of its maximum antenna gain, i.e., $\phi = 0^o$. Then, $\mathbf{F}(\Gamma)$ is given by (35).

The CRLBs $\kappa(\tau)$ and $\kappa(\Psi)$ for τ and Ψ respectively are the diagonal elements of the CRLB matrix $\kappa(\Gamma)$ where $\kappa(\Gamma) = (\mathbf{F}(\Gamma))^{-1}$. The position error bound (PEB) is expressed as the square root of the addition of diagonal elements of $\kappa(\Gamma)$ as [31]:

$$\text{PEB} = \sqrt{[(\mathbf{F}(\Gamma))^{-1}]_{1,1} + [(\mathbf{F}(\Gamma))^{-1}]_{2,2}} = \sqrt{\kappa(\Psi) + \kappa(\tau)} \tag{36}$$

where

$$\kappa(\Psi) = \frac{N_0 B/(\mathcal{G}P_t)}{(\frac{2\pi\bar{d}}{\lambda})^2 cos^2(\Psi)\beta^2\left[\sum_{n_t=0}^{N-1} n_t^2 - \left(\sum_{n_t=0}^{N-1} n_t\right)^2/N\right]}$$

$$\kappa(\tau) = \frac{N_0 B}{\mathcal{G}P_t}\frac{N\sum_{n_t=0}^{N-1} n_t^2}{(2\pi f_c)^2\beta^2\left[\sum_{n_t=0}^{N-1} n_t^2 - \left(\sum_{n_t=0}^{N-1} n_t\right)^2/N\right]}. \tag{37}$$

Since $1/f_c \ll 1$, CRLB $\kappa(\tau) \ll \kappa(\Psi)$. Hence, for $\bar{d} = \lambda/2$, we have

$$\text{PEB} \approx \sqrt{\kappa(\Psi)} = \frac{\sqrt{N_0 B/(\mathcal{G}P_t)}}{\frac{h\lambda}{4d_0}(\frac{d_0}{d})^{l/2} cos(\Psi)\left[\sum_{n_t=0}^{N-1} n_t^2 - \left(\sum_{n_t=0}^{N-1} n_t\right)^2/N\right]^{1/2}}. \tag{38}$$

Let the error in EAOA is $\psi = (\Psi - \Psi_e)$ (Fig. 2). Upper bound on the variance of ψ is quite complex to derive. Nevertheless, we can find the distribution of ψ using PEB assuming very low SNR at cell edge in the worst case, as stated below.

Lemma 3: The distribution of localization error ψ of an UE having EAOA Ψ and HPBW θ with respect to s^{th} sector is given as:

$$\text{Prob}(|\psi| \leq \epsilon) = \left[1 - \frac{\kappa(\Psi)}{\epsilon^2}\right]. \tag{39}$$

Proof. Using Chebyshev's inequality we have

$$\text{Prob}(|\psi| \geq \epsilon) \leq \left[\frac{\text{variance}(\psi)}{\epsilon^2}\right] = \frac{\kappa(\Psi)}{\epsilon^2} \quad \text{as the CRLB.}$$

$$\Rightarrow \text{Prob}(|\psi| \leq \epsilon) = \left[1 - \frac{\kappa(\Psi)}{\epsilon^2}\right], \quad \text{for} \quad -\frac{\theta}{2} \leq \Psi \leq \frac{\theta}{2}.$$

□

C. Complexity analysis

1) Subcarrier and power allocation

The joint subcarrier–power allocation with minimum UE rate constraint problem $\mathcal{P}1$ is a mixed integer NP-hard problem. For a system with K_s UEs and N_c subcarriers, $\mathcal{P}1$ has $K_s N_c$ variables, atleast $K_s + N_c$ constraints, total UE–subcarrier $K_s^{N_c}$ combinations, and $K_s^{N_c}$ comparisons. For power allocation per UE–subcarrier combination, complexity is $\mathcal{O}(K_s N_c \log_2 N_c)$. Thus, the complexity of solving $\mathcal{P}1$ using exhaustive search is $\mathcal{O}(K_s^{N_c}(K_s N_c \log_2 N_c))$.

On the other hand, sub-optimum approach (Steps (i) through (iv)) is computationally less complex. It requires $K_s N_c$ comparisons in Step (ii), at most $(K_s - 1)N_c$ multiplications and $(K_s - 1)N_c$ summations in Step (iii) to compute cost functions, approximately $2N_c^2$ comparisons for subcarrier reallocation, and in the worst case $N_c \log_2 N_c$ power allocation computations in Step (iv). Hence, the overall complexity of sub-optimal approach is reduced to $\mathcal{O}(N_c^2)$.

2) Sectorization

To find the average long-run UE rate with total number of sectors S in the worst case scenario of all sectors being active, approximately $S(3K_s N_c + 2N_c^2 + N_c \log_2 N_c)$ computations are required. Therefore, to find S^* in fixed time scheduling scheme an exhaustive search over $\forall S \in \mathcal{S}$ is performed. As a result, the worst case complexity to estimate S^* is $\mathcal{O}(|\mathcal{S}|SN_c^2)$, where

$$\mathbf{F}(\Gamma) = \frac{\mathcal{G}P_t}{N_0 B}\begin{bmatrix} (2\pi f_c)^2\beta^2 N & (2\pi f_c)(\frac{2\pi\bar{d}}{\lambda})\beta^2 cos(\Psi)\sum_{n_t=0}^{N-1} n_t \\ (2\pi f_c)(\frac{2\pi\bar{d}}{\lambda})\beta^2 cos(\Psi)\sum_{n_t=0}^{N-1} n_t & (\frac{2\pi\bar{d}}{\lambda})^2 cos^2(\Psi)\beta^2\sum_{n_t=0}^{N-1} n_t^2 \end{bmatrix} \tag{35}$$

S can have the maximum value of 360 for $\theta = 1^o$. In variable time scheduling, if η_{itr} iterations are required to find α, then the worst case complexity is $\mathcal{O}(\eta_{itr}|\mathcal{S}|SN_c^2)$.

3) Reduced-complexity method

In reduced complexity method, subcarrier allocation is the same as in $\mathcal{P}1$, but it is computed for only one sector per sectorization scheme to form the dictionary at BS, which is updated only when channel statistics change. Accordingly, the complexity reduces to $\mathcal{O}(|\mathcal{S}|N_c^2)$. Further, to find $\hat{S}^*$ for a given UE population, maximum $|\mathcal{S}|N_c$ comparisons are needed.

IV. Simulation Results and Discussion

In this section, we present the numerical simulation results generated using MATLAB with varying system configurations. For the simulation set-up, we consider a system operating at 60 GHz carrier frequency with $B = 1$ GHz bandwidth, and $N_c = 30$ subcarriers. We vary the beamwidth in the range $1^o \leq \theta \leq 30^o$ in steps of 1^o. Various other channel and system parameters considered are: Rician fading parameter $= 8$ dB [32], EIRP $= 52$ dBm, $G_0 = 5$ dBi, cell radius $= 200$ m, $R_0 = 10$ Mbps, noise spectral density $N_0 = -174$ dBm/Hz, $\zeta_{LOS} = 15, \zeta_{NLOS} = 45$, and outage SNR threshold $= -6$ dB. We execute the simulations broadly for two cases: firstly considering no localization error for the two scheduling schemes, namely, fixed time and variable time scheduling, and secondly considering localization error with fixed time scheduling.

A. Optimal sectorization scheme with perfect UE location information

1) Sub-optimum resource allocation per sector

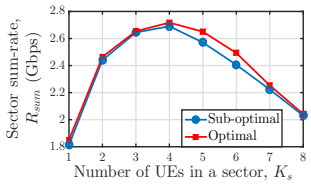

Figure 3. Sector sum rate versus number of UEs plot for optimal and sub-optimal resource allocation techniques.

Due to the NP-hard nature of $\mathcal{P}1$, per-sector resource allocation is performed in a sub-optimal way that maximizes the sector's instantaneous sum rate while satisfying minimum rate constraints. Fig. 3 compares per-sector sum rate achieved using optimal resource allocation and using sub-optimal method for different values of K_s, using Monte-Carlo simulations. As the computation time for a single Monte-Carlo iteration in optimal resource allocation is of order $K_s^{N_c}$, to perform simulations in this case we took $N_c = 8$ and $B = 250$ MHz while keeping the other parameters same as described above. Form Fig. 3 we observe that both resource allocation schemes show a similar trend on varying K_s, where the optimal solution gives an upper bound.

2) Fixed time scheduling

With decreasing sector beamwidth, the same frequency band can be used in the cell more often, thereby increasing the spatial reuse factor. For a fixed UE population, although a narrower beam provides the benefit of increased cell spatial reuse factor, Fig. 4(a) shows that a narrow beam does not necessarily result in maximization of the average long-run UE rate $\bar{R}$. The decrease in $\bar{R}$ is due to two reasons: (i) with a beam serving very few UEs, the channel utilization is not efficient (*Remark* 2), and (ii) total delay between the next schedule of each sector increases with more cell partitions.

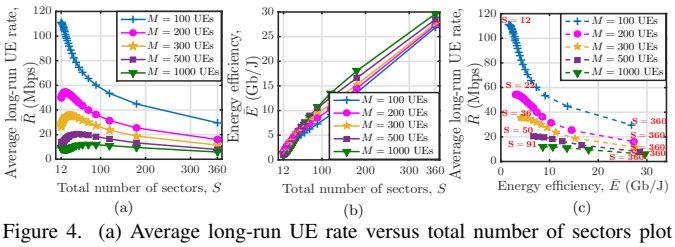

Figure 4. (a) Average long-run UE rate versus total number of sectors plot for uniformly distributed UE population; (b) BS energy efficiency versus total number of sectors plot for uniformly distributed UE population. (c) Pareto front for average long-run UE rate and BS energy efficiency optimization.

Suppose S_R^* is the optimal sectorization scheme which offers maximum $\bar{R}$ without BS energy optimization constraint. Fig. 4(a) shows that there exists an S_R^* corresponding to $\theta_{min} \leq \theta^* \leq \theta_{max}$, for which $\bar{R}$ is maximized. Moreover, the optimal sector beamwidth gets narrower with increasing UE density. Fig. 4(b) demonstrates that $\bar{E}$ is nearly proportional to cell sectorization, which is because a higher spatial reuse factor increases the cell sum-throughput with minimal increase in transmit energy requirement for one complete cell sweep duration. From combined observations of Fig. 4(a) and 4(b), we conclude that for $S \geq S_R^*$ a Pareto front exists between $\bar{E}$ and $\bar{R}$ as captured in Fig. 4(c). For instance, for $M = 200$ with corresponding $S_R^* = 22$, a Pareto optimal condition between $\bar{R}$ and $\bar{E}$ is observed in Fig. 4(c) over the range of $S = [22, 360]$. Thus, the sectorization schemes $S < S_R^*$ are not optimal either from $\bar{R}$ or from $\bar{E}$ point of view.

We further compare $\bar{R}$ and $\bar{E}$ performances of the single beam system with multiple analog beams system (MABS) and JSDM [15], as shown in Fig. 5. In MABS we consider N_{RF} RF chains at BS, each connected to N/N_{RF} antenna elements. Each beam (one from each RF chain) serves a fraction of the total number of sectors. The beams are fanned out to have maximum possible spatial separation to minimize inter-beam interference. The hardware power consumption of a mmWave system is given as $P_{NRF}(N_{RF}) = P_{PA} + P_{LO} + N_{RF}(2P_{DAC} + 2P_{RF})$, where P_{DAC}, P_{RF}, P_{LO}, and P_{PA} represent the power consumption by the digital-to-analog

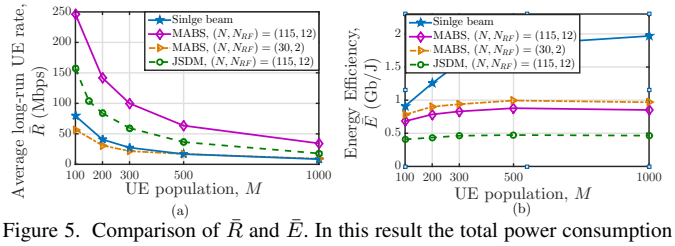

Figure 5. Comparison of $\bar{R}$ and $\bar{E}$. In this result the total power consumption is given as $P_{TOTAL} = P_{NRF} + N_{RF}P_t$ and consider the practical gain pattern given by (28). Simulation parameters for single-beam system are $(N, N_{RF}) = (\lceil 2/\theta^* \rceil, 1)$; for JSDM $(N, N_{RF}) = (115, 12)$; for MABS $(N, N_{RF}) = (115, 12)$ and $(N, N_{RF}) = (30, 2)$.

converter, the RF chain, the local oscillator, and the power amplifier, respectively [33].

In MABS simulations, for a given value of N, we consider $N_{RF} = \text{argmax}\ \bar{E}(N_{RF})$. From Figs. 5(a) and 5(b) it is observed that single beam is a better choice over MABS with $(N, N_{RF}) = (30, 2)$, in terms of $\bar{R}$ as well as $\bar{E}$. However, with $(N, N_{RF}) = (115, 12)$ MABS offers improved $\bar{R}$ at the cost of decreased $\bar{E}$. Therefore, proper selection of (N, N_{RF}) is crucial in MABS. Another notable point is that MABS performs better than JSDM for the same number of N_{RF} units because it better utilizes the wideband channel by supporting multiple UEs over different subcarriers of OFDM symbol whereas in JSDM each UE has a dedicated RF chain and full channel resources assigned to it. The detailed performance analysis of MABS incorporating inter-beam interference, and $\bar{R}$ and $\bar{E}$ trade-off with respect to (N, N_{RF}) is left as future work.

Reduced complexity method: For $\bar{R}$ maximization, as discussed in Section III-A2, we estimate the sub-optimal sectorization scheme $\hat{S}^*$, at which maximum $\bar{R}$ is achieved, with a considerably reduced number of computations. For this, we generate the dictionary $\mathcal{R}_{sum}$ at the BS using Monte-Carlo simulations. Fig. 6(a) shows the dependency of sector sum rate on K_s and S. Fig. 6(b) shows comparison of

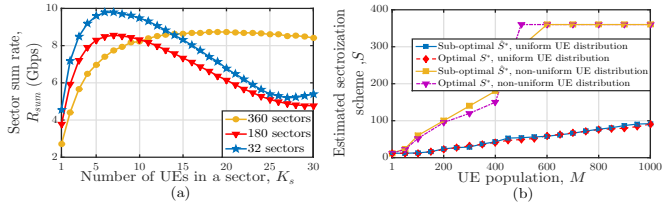

Figure 6. (a) Plot of sum rate per sector versus number of UEs in the sector for three different sectorization schemes. (b) Comparison of optimal S^* obtained through exhaustive search method and sub-optimal $\hat{S}^*$ obtained using reduced complexity search method for uniform and non-uniform UE distribution.

the sub-optimal $\hat{S}^*$ with the optimal S^*. We observe that with the reduced complexity method $\hat{S}^*$ is nearly affine with UE population (below $M = 50$, $S = 12$, since maximum beamwidth possible is 30^o) for $\theta_{min} \leq \theta^* \leq \theta_{max}$. Moreover, $\hat{S}^*$ is quite close to S^*. There is an error in estimation $|S^* - \hat{S}^*|$ because we construct the dictionary R_{sum} by using statistical UE and channel distribution, and approximation in (24), whereas in estimation of S^* we use UE instantaneous channel state information. Thus, we can get quite close results with much reduced complexity as compared to exhaustive search method, and the update in the dictionary is required only when statistical channel parameters change.

3) Variable time scheduling

We simulate non-uniform UE distribution by dividing the cell into 6 zones, i.e., $Z = 6$ with p_z vector for the zones being $[0.9, 0.05, 0.03, 0.01, 0.008, 0.002]$. As described in Section III-A3, we consider fixed time scheduling scheme as a benchmark for variable time scheduling scheme. Thereafter, we use Algorithm 1 and Algorithm 2 to find the optimal shaping parameter α_R and α_E, respectively, for improved performance over fixed scheduling. Fig. 7(a) illustrates that there exits a non-zero α_R that offers a higher value of average long-run UE

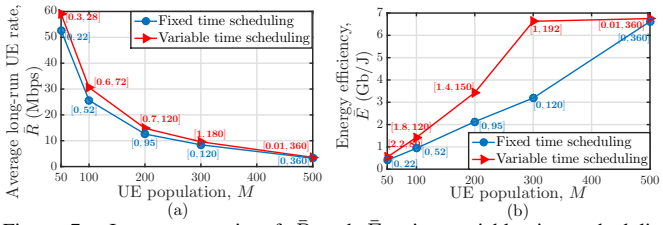

Figure 7. Improvement in of $\bar{R}$ and $\bar{E}$ using variable time scheduling scheme over fixed time scheduling scheme. (a) shows the improvement in maximum achievable average long-run UE rate for different UE population, using variable time scheduling scheme with $[\alpha = \alpha_R, S = S'^*]$ over fixed time scheduling scheme at $[\alpha = \alpha_0, S = S^*]$; and (b) shows the improvement in maximum achievable BS energy efficiency using variable time scheduling scheme with $[\alpha = \alpha_E, S = S''^*]$ over fixed time scheduling scheme at $[\alpha = \alpha_0, S = S^*]$.

rate $\bar{R}_{\alpha_R}(S'^*)$ compared to fixed scheduling scheme $\bar{R}_{\alpha_0}(S^*)$, such that $\bar{E}_{\alpha_R}(S'^*) \geq \bar{E}_{\alpha_0}(S^*)$ and $J_{\alpha_R}(S'^*) \geq J_{\alpha_0}(S^*)$ (as shown in Fig. 7(a)).

Similarly, a non-zero α_E exists that offers an increased energy efficiency attainable $\bar{E}_{\alpha_E}(S''^*)$ as compared to energy efficiency achieved using fixed scheduling scheme $\bar{E}_{\alpha_0}(S^*)$, such that $\bar{R}_{\alpha_E}(S''^*) \geq \bar{R}_{\alpha_0}(S^*)$ and $J_{\alpha_E}(S''^*) \geq J_{\alpha_0}(S^*)$ (as shown in Fig. 7(b)). We also note from the simulation results that $\alpha_E \geq \alpha_R$. For example, for $M = 200$ non-uniformly distributed UEs, the fixed scheduling scheme has $S^* = 95, \alpha = 0$, $\bar{R}_{\alpha_0}(S^*) = 12.58$ Mbps, and $\bar{E}_{\alpha_0}(S^*) = 2.13$ Gb/J. With variable time scheduling the performance is improved. At the value of α equal to $\alpha_R = 0.7$ a value of $\bar{R}_{\alpha_R}(S'^*) = 14.79$ Mbps is achieved at $S'^* = 120$. If the shaping parameter is set to $\alpha = \alpha_E = 1.4$ with $S''^* = 150$, we get gain in BS energy efficiency, i.e, $\bar{E}_{\alpha_E}(S''^*) = 3.44$ Gb/J.

B. Optimal sectorization scheme with estimated UE position information

Due to localization error some UEs that are wrongly detected to be in another sector will experience different EAOA (increased EAOA). Since in practical scenario beam gain has a fan shaped pattern given by (28), gain falls sharply outside HPBW and the UEs lying in this area will be the main attribute for the decrease in average long-run UE rate $\bar{R}$ (if any). Thus, the effect of localization error on the $\bar{R}$ is mainly attributed by the UEs that are wrongly detected to be in another sector.

In the simulations, we generate the localization error for each UE using the distribution given by (39) and the antenna gain at Ψ EAOA using (28). Since we have the knowledge of only lower bounds on variance of the estimator, to study worst-case scenario we carry out simulations assuming $\mathcal{G} = 2$ and minimum SNR requirement at cell edge to be -6 dB. The result in Fig. 8(a) illustrates a decrease in PEB with increasing

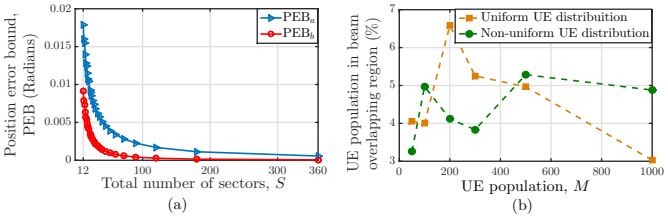

Figure 8. (a) Position error bound of Ψ_e versus number of sectors in worst-case scenario. PEB_a is the PEB obtained using (38); PEB_b is the PEB achieved in [5]. (b) Percentage of UEs lying in the beam overlapping region due to localization error.

sector partitions. The PEB obtained from our simulation is of increased magnitude as compared to the work in [5]. The reason being that the increase in SNR (due to enhanced beam directivity) is nullified by reduced transmit power, thereby keeping SNR invariant of beamwidth.

We compare the simulation results of three cases: (a) ideal case of resource allocation without localization error and constant beam gain within the sector ($G = NG_0$), (b) resource allocation with no localization error but angle-dependent beam gain (given by (28)), and (c) resource allocation with localization error (given by (39)) along with angle-dependent beam gain (given by (28)). Fig. 9(a) and 9(b) show that, though $\bar{R}$ in case (b) is of reduced magnitude as compared to $\bar{R}$ in case (a), the nature of $\bar{R}$ as a function of S is same for both. Moreover, in case (c) $\bar{R}$ further degrades slightly with localization error. However, S^* remains invariant.

From the results in Fig. 4(a) we noted that, on increasing M, θ^* shifts to a smaller value; thus K_s in a sector is such that it maximizes the sector sum rate. Moreover, our insight is that with reducing HPBW the beam overlapping region shrinks. We omit the related mathematics for brevity. Consequently, the percentage of UEs falling in the overlapping region due to localization error is very small. To verify this claim in Fig. 8(b) we plotted the percentage of UEs falling in beam overlapping area for different sets of $(M, S^*(M))$. The figure verifies that the population in the overlapping region due to erroneous position estimation is less than 10% even for M as high as 1000. Therefore, the optimal number of sectors remains unchanged. Similar behaviour is observed with non-uniformly distributed UEs as well.

Further, from Fig. 9(a) and 9(b) we observe that the graphs have convex behaviour till some S, after which they show concavity. This is explained as follows: (i) When $\{S \in \mathcal{S}|K_s \gg N_c\}$, the subcarriers in a sector are assigned to only N_c out of K_s UEs that have best channel conditions, such that the minimum rate requirement is fulfilled for only N_c UEs. In addition to this, the cell sweep time is of very short duration for small S. (ii) When $\{S \in \mathcal{S}|K_s > N_c\}$, the subcarriers are still assigned to only N_c out of K_s UEs having best channel conditions. However, with decreasing domain size K_s, the chances of the majority of UEs having good channel conditions reduces. Further sweep time also increases with S. As a result, it decreases the average long-run UE rate. (iii) When $\{S \in \mathcal{S}|K_s \leq N_c\}$, all the UEs are allocated subcarrier such that the total sum rate is maximized with a minimum rate guarantee to all UEs. Although cell sweep time is still increasing, the gain in the instantaneous UEs rate outweighs the increased sweep duration effect. Therefore, the sector sum rate begins to improve, resulting in the concave nature of the average long-run UE rate curve.

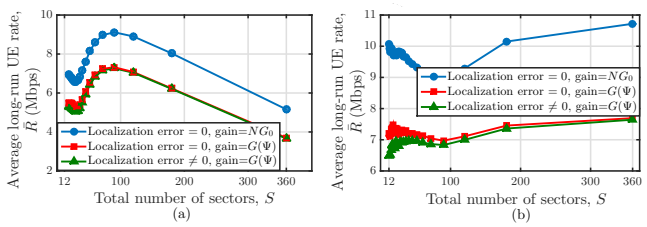

Figure 9. Average long-run UE rate versus total number of sectors plot for (a) 1000 UEs uniformly distributed in the cell, and (b) 500 UEs non-uniformly distributed in the cell with probability distribution vector $[0.9, 0.05, 0.03, 0.01, 0.008, 0.002]$.

V. Conclusion

This paper has focused on the performance of cellular communication at wideband mmWave frequencies serving UEs using a single analog beam from one RF chain in conjunction with scheduling operation such that radiation levels in the environment are maintained constant. The proposed system employs a sectored-cell spatial multiplexing approach and aims to serve more than one UE during the beam sojourn time. It has been noted that, although a narrow beam provides increased channel spatial reuse, it does not necessarily result in an increased average long-run UE data rate. Therefore, to cater to a specific UE population, an optimal sector beamwidth is required that maximizes the average long-run UE rate and improves QoS such that the BS transmit power requirement is also reduced. From the BS perspective, it is always optimal to use the narrowest possible beamwidth. Hence, a Pareto front exists on average long-run UE rate maximization and BS energy efficiency. Furthermore, the proposed reduced-complexity method has provided an insight that channel utilization is maximum when it has an optimum number of UEs to serve over the wideband channel. Therefore, with increasing UE population, if the number of sectors in a cell is maintained such that each sector operates at its optimum channel usage, then the highest average long-run UE rate can be achieved.

Further, it has been observed that, in the case of non-uniformly distributed UEs, the variable time scheduling scheme can further improve the average long-run UE rate and BS energy efficiency as compared to fixed time scheduling. The results have shown that non-uniform UE distribution corresponds to shifting to a narrower beam with the increasing variance of UE distribution. Further, studies by accounting for the localization uncertainty have revealed that, in a static UE scenario, though localization error reduces average long-run UE data rate, optimal sector beamwidth selection remains unaffected, even for non-uniformly distributed UEs. The analysis of single analog beam with sectored-cell multiplexing is extendible to the case of multiple concurrent beams serving the cell. However, it would require an equal number of RF chains and complex inter-beam interference analysis. In conclusion, although mmWave holds the potential to extend the limits of current cellular communication because of large spectrum availability, realizing these potential requires redesigning system model and architecture.

Appendix

A. Proof of Lemma 1

Lemma 1: For a fixed α and $K_s \leq N_c$, Jain's fairness index J is an increasing function of S, when resource allocation is performed using sum rate maximization approach.

Proof. Suppose there are $1 < m \leq N_c$ UEs in s^{th} sector, such that the instantaneous rate of UEs is $r_1, r_2, \cdots, r_m$. Let

$\alpha = 0$ and $t_s = 1$ unit. Then by (19), we have

$$J(S) = \frac{1}{m}\frac{\left(\frac{r_1}{S} + \cdots + \frac{r_m}{S}\right)^2}{\left(\left(\frac{r_1}{S}\right)^2 + \cdots + \left(\frac{r_m}{S}\right)^2\right)} = \frac{1}{m}\frac{(r_1 + \cdots + r_m)^2}{({r_1}^2 + \cdots + {r_m}^2)}$$
$$= \frac{1}{m}\frac{x^2}{y} \text{ (say).}$$

Now suppose the sectorization scheme is changed to $\tilde{S} = S + 1$ with UEs position in the cell unchanged. This results in redistribution of UEs among the sectors. With increasing number of sectors the degree of spatial multiplexing increases, leading to increased instantaneous sum rate of the UEs. Let us consider that due to redistribution one of the UEs falls into another sector while other $m-1$ UEs still remain together in one sector. Let the increment in instantaneous rate of the m^{th} UE falling into another sector be δ. Therefore,

$$J(\tilde{S}) = \frac{1}{m}\frac{(r_1 + \cdots + r_m + \delta)^2}{(r_1^2 + \cdots + (r_m + \delta)^2)} = \frac{1}{m}\frac{(x+\delta)^2}{(y + \delta^2 + 2\delta r_m)}$$
$$\geq \frac{1}{m}\frac{(x+\delta)^2 - (\delta^2 + 2\delta r_m)}{(y + \delta^2 + 2\delta r_m) - (\delta^2 + 2\delta r_m)}$$
$$\text{(since } 0 \leq J(S) \leq 1\text{)}$$
$$= \frac{1}{m}\frac{(x^2 + 2\delta(r_1 + \cdots + r_{m-1}))}{y} \geq \frac{1}{m}\frac{x^2}{y}.$$

Hence, $J(\tilde{S}) \geq J(S)$, i.e, $J(S)$ is an increasing function of S. □

B. Derivation of array gain

$$G(\hat{\theta}) = N\left|\left[\frac{1}{N}\frac{\sin(\frac{N\pi}{2}\sin(\hat{\theta}))}{\sin(\frac{\pi}{2}\sin(\hat{\theta}))}\right]\right|^2, \text{ for } -\pi \leq \hat{\theta} \leq \pi \quad \text{(A.1)}$$

Proof. Gain of an broadside ULA, with 100% efficiency, with respect to angle of departure $\hat{\theta}$ of a signal from direction of maximum array gain is given as [29]:

$$G(\hat{\theta}) = \frac{4\pi|AF(\hat{\theta})|^2}{\int_0^{2\pi}\int_0^{\pi}|AF(\hat{\theta})|^2 \sin\hat{\theta}\ \mathrm{d}\hat{\theta}\ \mathrm{d}\hat{\phi}}$$

where $AF(\hat{\theta})$ is the normalized antenna array factor. For zero phase difference excitation and half wavelength antenna spacing, it is given as [29]:

$$AF(\hat{\theta}) = |\mathbf{a}(\hat{\theta})|^2 = \frac{1}{N}\left[\frac{\sin(\frac{N\pi}{2}\sin\hat{\theta})}{\sin(\frac{\pi}{2}\sin\hat{\theta})}\right].$$

Let
$$\mathcal{V} = \int_0^{\pi}|AF(\hat{\theta})|^2 \sin\hat{\theta}\ \mathrm{d}\hat{\theta}$$
$$= \int_0^{\pi}\left|\frac{\sin(\frac{N\pi}{2}\sin\hat{\theta})}{N\sin(\frac{\pi}{2}\sin\hat{\theta})}\right|^2 \sin\hat{\theta}\ \mathrm{d}\hat{\theta}.$$

Suppose $\omega = 0.5\pi\cos\theta$, then we have

$$\mathcal{V} = \frac{2}{N^2\pi}\int_{-\pi/2}^{\pi/2}\left|\frac{\sin(N\omega)}{\sin(\omega)}\right|^2 \mathrm{d}\omega. \quad \text{(A.2)}$$

Since, $\frac{\sin\ (N\omega)}{\sin\ (\omega)} = \frac{e^{jN\omega} - e^{-jN\omega}}{e^{j\omega} - e^{-j\omega}} = \frac{e^{j\omega}}{e^{jN\omega}}\frac{e^{2jN\omega} - 1}{e^{2j\omega} - 1}$

$$= e^{-j(N-1)\omega}\sum_{k=0}^{N-1} e^{2jk\omega}.$$

By Fourier series expansion,

$$\sum_{k=-\infty}^{+\infty} a_k e^{jk\omega} = \frac{\sin\ (N\omega)}{\sin\ (\omega)} = e^{-j(N-1)\omega}\sum_{k=0}^{N-1} e^{2jk}\omega$$

with only N non-zero coefficient of unity magnitude. Therefore, by Parseval's Identity we have

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}|f(x)|^2\ \mathrm{d}x = \sum_{k=-\infty}^{\infty}|a_k|^2 = N.$$

Hence, (A.2) can be solved as: $\mathcal{V} = (2/N^2\pi)(1/2)\int_{-\pi}^{\pi}\left|\frac{\sin N\omega}{\sin\hat{\phi}}\right|^2 \mathrm{d}\omega = 2/N$. Therefore, in general, the array gain at an angle $\hat{\theta}$ from the direction of maximum gain is:

$$G(\hat{\theta}) = \frac{4\pi|AF(\hat{\theta})|^2}{\frac{2}{N}\int_0^{2\pi}\mathrm{d}\hat{\phi}} = N\left|AF(\hat{\theta})\right|^2$$
$$= N\left|\left[\frac{1}{N}\frac{\sin(\frac{N\pi}{2}\sin(\hat{\theta}))}{\sin(\frac{\pi}{2}\sin(\hat{\theta}))}\right]\right|^2, \text{for } -\pi \leq \hat{\theta} \leq \pi.$$

□

C. Proof of Lemma 2

Lemma 2: The FIM for the received signal is given as:

$$\mathbf{F}(\Gamma) = \frac{1}{N_0 B}\sum_{g=1}^{\mathcal{G}}\mathrm{Re}(\nabla y^g \nabla^H y^g). \quad \text{(A.3)}$$

Proof. From CRLB definition [30], we have

$$\mathrm{var}(\hat{\Gamma} - \Gamma) \geq -\left[\mathbb{E}_{\mathbf{Y}|\Gamma}\left(\frac{\partial^2 \log f(\mathbf{Y}|\Gamma)}{\partial\Gamma\partial\Gamma^T}\right)\right]^{-1} \equiv (\mathbf{F}(\Gamma))^{-1}$$

where $\mathbf{Y} = [y^1, \ldots, y^{\mathcal{G}}]^T$. Let $w(\Gamma) = \sqrt{N}\beta e^{-j2\pi f_c \tau}\mathbf{a}^H(\Psi)\mathbf{s}$. Then, for g^{th} transmission (31) can be written as $y^g = w^g(\Gamma) + \nu^g$. Hence, the log-likelihood function is:

$$f(\mathbf{Y}|\Gamma) = \frac{1}{\sqrt{2\pi\sigma^2}}\left[e^{-\frac{1}{2\sigma^2}(y^1 - w^1(\Gamma))}, \ldots, e^{-\frac{1}{2\sigma^2}(y^{\mathcal{G}} - w^{\mathcal{G}}(\Gamma))}\right]^T$$
$$\Rightarrow \log f(\mathbf{Y}|\Gamma) = \left[-\frac{1}{2}\log 2\pi\sigma^2 - \frac{1}{2\sigma^2}(y^1 - w^1(\Gamma)),\right.$$
$$\left. \ldots, -\frac{1}{2}\log 2\pi\sigma^2 - \frac{1}{2\sigma^2}(y^{\mathcal{G}} - w^{\mathcal{G}}(\Gamma))\right]^T.$$

Taking the first derivative, we have:

$$\frac{\partial f(\mathbf{Y}|\Gamma)}{\partial\Gamma} = -\frac{1}{2\sigma^2}\left[\nabla y^1, \ldots, \nabla y^{\mathcal{G}}\right]^T.$$

Taking second derivative and negative expected value over Γ, we get:

$$-\mathbb{E}_{\mathbf{Y}|\Gamma}\left(\frac{\partial^2 f(\mathbf{Y}|\Gamma)}{\partial\Gamma\partial\Gamma^T}\right) = \frac{1}{2\sigma^2}\sum_{g=1}^{\mathcal{G}}\mathrm{Re}(\nabla y^g \nabla^H y^g)$$
$$\Rightarrow \mathbf{F}(\Gamma) = \frac{1}{N_0 B}\sum_{g=1}^{\mathcal{G}}\mathrm{Re}(\nabla y^g \nabla^H y^g).$$

□

REFERENCES

[1] I. A. Hemadeh, K. Satyanarayana, M. El-Hajjar, and L. Hanzo, "Millimeter-wave communications: Physical channel models, design considerations, antenna constructions, and link-budget," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 2, pp. 870–913, 2017.

[2] X. Wang, L. Kong, F. Kong, F. Qiu, M. Xia, S. Arnon, and G. Chen, "Millimeter wave communication: A comprehensive survey," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 3, pp. 1616–1653, 2018.

[3] S. Kutty and D. Sen, "Beamforming for millimeter wave communications: An inclusive survey," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 2, pp. 949–973, 2015.

[4] T. S. Rappaport, J. H. Reed, and B. D. Woerner, "Position location using wireless communications on highways of the future," *IEEE Commun. Mag.*, vol. 34, no. 10, pp. 33–41, 1996.

[5] A. Fascista, A. Coluccia, H. Wymeersch, and G. Seco-Granados, "Millimeter-wave downlink positioning with a single-antenna receiver," *IEEE Trans. Wireless Commun.*, vol. 18, no. 9, pp. 4479–4490, 2019.

[6] M. Koivisto, M. Costa, J. Werner, K. Heiska, J. Talvitie, K. Leppänen, V. Koivunen, and M. Valkama, "Joint device positioning and clock synchronization in 5g ultra-dense networks," *IEEE Trans. Wireless Commun.*, vol. 16, no. 5, pp. 2866–2881, 2017.

[7] I. Ahmed, H. Khammari, A. Shahid, A. Musa, K. S. Kim, E. De Poorter, and I. Moerman, "A survey on hybrid beamforming techniques in 5G: Architecture and system model perspectives," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 4, pp. 3060–3097, 2018.

[8] J. Lee, G.-T. Gil, and Y. H. Lee, "Channel estimation via orthogonal matching pursuit for hybrid MIMO systems in millimeter wave communications," *IEEE Trans. Commun.*, vol. 64, no. 6, pp. 2370–2386, 2016.

[9] H. Shokri-Ghadikolaei and C. Fischione, "The transitional behavior of interference in millimeter wave networks and its impact on medium access control," *IEEE Trans. Commun.*, vol. 64, no. 2, pp. 723–740, 2015.

[10] 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.

[11] Q. Xue, X. Fang, M. Xiao, and L. Yan, "Multiuser millimeter wave communications with nonorthogonal beams," *IEEE Trans. Veh. Technol.*, vol. 66, no. 7, pp. 5675–5688, 2016.

[12] 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, 2015.

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

[14] M. Nair, Q. Z. Ahmed, J. Wang, and H. Zhu, "Low-complexity hybrid digital-to-analog beamforming for millimeter-wave systems with high user density," in *2017 IEEE 85th IEEE Veh. Technol. Conf. (VTC Spring)*, 2017, pp. 1–5.

[15] 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.

[16] N. Eshraghi, V. Shah-Mansouri, and B. Maham, "Fair beamwidth selection and resource allocation for indoor millimeter-wave networks," in *2017 IEEE Int. Conf. Commun. (ICC)*, 2017, pp. 1–6.

[17] S. Doğan, M. Karabacak, and H. Arslan, "Optimization of antenna beamwidth under blockage impact in millimeter-wave bands," in *2018 IEEE 29th Int. Symp. Personal, Indoor and Mobile Radio Commun. (PIMRC)*, 2018, pp. 1–5.

[18] FCC, 2015. [Online]. Available: https://apps.fcc.gov/edocs_public/attachmatch/FCC-15-138A1.pdf

[19] 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.

[20] D. J. Daley and D. Vere-Jones, *An introduction to the theory of point processes: volume II: general theory and structure*. Springer Science & Business Media, 2007.

[21] S. Sun, T. A. Thomas, T. S. Rappaport, H. Nguyen, I. Z. Kovacs, and I. Rodriguez, "Path loss, shadow fading, and line-of-sight probability models for 5G urban macro-cellular scenarios," in *2015 IEEE Globecom Workshops*, 2015, pp. 1–7.

[22] 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.

[23] J. Lee, J. Liang, J.-J. Park, and M.-D. Kim, "Beamwidth-dependent directional propagation loss analysis based on 28 and 38 GHz urban micro-cellular (UMi) measurements," in *2017 IEEE 86th Veh. Technol. Conf. (VTC-Fall)*, 2017, pp. 1–5.

[24] W. Yu and R. Lui, "Dual methods for nonconvex spectrum optimization of multicarrier systems," *IEEE Trans. Commun.*, vol. 54, no. 7, pp. 1310–1322, 2006.

[25] S. Boyd, S. P. Boyd, and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.

[26] Y. J. Zhang and K. B. Letaief, "Multiuser adaptive subcarrier-and-bit allocation with adaptive cell selection for OFDM systems," *IEEE Trans. Wireless Commun.*, vol. 3, no. 5, pp. 1566–1575, 2004.

[27] A. B. Sediq, R. H. Gohary, and H. Yanikomeroglu, "Optimal tradeoff between efficiency and Jain's fairness index in resource allocation," in *2012 IEEE 23rd Int. Symp. Personal, Indoor and Mobile Radio Commun. (PIMRC)*, 2012, pp. 577–583.

[28] C. Jiang, Y. Shi, Y. T. Hou, and S. Kompella, "On optimal throughput-energy curve for multi-hop wireless networks," in *2011 Proc. IEEE INFOCOM*, 2011, pp. 1341–1349.

[29] C. A. Balanis, *Antenna theory: analysis and design*. John wiley & sons, 1997.

[30] M. S. Kay, *Fundamentals of statistical signal processing*. Prentice Hall PTR, 1993.

[31] D. Dardari, E. Falletti, and M. Luise, *Satellite and terrestrial radio positioning techniques: A signal processing perspective*. Academic Press, 2012.

[32] T. Zhou, C. Tao, L. Liu, and Z. Tan, "Ricean K-factor measurements and analysis for wideband radio channels in high-speed railway U-shape cutting scenarios," in *2014 IEEE 79th Veh. Technol. Conf. (VTC Spring)*, 2014, pp. 1–5.

[33] M. Shehata, A. Mokh, M. Crussière, M. Hélard, and P. Pajusco, "On the equivalence of hybrid beamforming to full digital zero forcing in mmWave MIMO," in *Proc. Int. Conf. Telecommun. (ICT)*, 2019, pp. 1–7.

Nancy Varshney received the B.Tech. degree in Electronics and Communication Engineering from Dr. A.P.J. Abdul Kalam Technical University, Lucknow, India, in 2015, and the M.Tech. degree from Indian Institute of Technology BHU, Varanasi, India, in 2018. She is currently pursuing the Ph.D. degree from Bharti School of Telecommunication, at Indian Institute of Technology Delhi, New Delhi, India. Her current research interests include beamforming, mmWave communications and networks.

Swades De [(S'02-M'04-SM'14)] is a Professor in the Department of Electrical Engineering at IIT Delhi. Before moving to IIT Delhi in 2007, he was a Tenure-Track Assistant Professor of Electrical and Computer Engineering at the New Jersey Institute of Technology (2004-2007). He worked as an ERCIM post-doctoral researcher at ISTI-CNR, Pisa, Italy (2004), and has nearly five years of industry experience in India on telecom hardware and software development (1993-1997, 1999). His research interests are broadly in communication networks, with emphasis on performance modeling and analysis. Current directions include energy harvesting sensor networks, broadband wireless access and routing, cognitive/white-space access networks, smart grid networks, and IoT communications. Dr. De currently serves as an Area Editor for the IEEE COMMUNICATIONS LETTERS and Elsevier Computer Communications, and an Associate Editor for the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, the IEEE WIRELESS COMMUNICATIONS LETTERS, and the IEEE NETWORKING LETTERS.