Position Aided Beam Learning for Initial Access in mmWave MIMO Cellular Networks

Anzhong Hu, *Member, IEEE* and Jiguang He, *Member, IEEE*

***Abstract*—In this paper, beam learning based on position information (PI) about mobile station positions in the initial access (IA) of millimeter wave (mmWave) multiple input multiple output cellular networks is investigated. The existing PI based IA procedure cannot efficiently tackle the position inaccuracy and blockage or may cause a long IA delay because of the inefficient beam learning. Based on the sparse scattering of mmWave signals, the serving area is partitioned into smaller areas and the beams are learned for each small area. Moreover, the number of learned beams is restricted and fixed after learning. Thus, the impact of position inaccuracy and blockage can be mostly mitigated and the IA delay is short in each successful IA. The analysis shows the lower bound of the probability of miss detection. Additionally, the simulation results show that the proposed approach can achieve a reasonable IA delay and superior IA performance than other PI based approaches.**

***Index Terms*—MIMO systems, millimeter wave, beam, context information.**

I. Introduction

As the frequency band below 6 GHz is mostly occupied by current communication systems, the frequency band above 6 GHz has been attracting attention in recent years. Moreover, the channel in millimeter wave (mmWave) band above 30 GHz has been measured and considered as a candidate for the next generation mobile communication system [1][2]. The mmWave band brings both a shortcoming and a benefit. On one hand, the propagation loss in mmWave band is much higher than that with frequency below 6 GHz. On the other hand, the short wave length facilitates the placement of a large number of antennas in a constrained space, which constitute the multiple-input multiple-output (MIMO) system and can provide a high array gain. Thus, beamforming is essential in mmWave MIMO systems as it can achieve a high array gain to compensate for the high propagation loss.

However, beamforming is challenging in initial access (IA) in mmWave MIMO cellular networks, where IA refers to the procedures in which a mobile station (MS) establishes a physical link with a base station (BS), i.e., from idle mode to connected mode [3][4]. Because the MS and the BS have no knowledge of the channel state information, they have to try all the beam directions and find the best one. As can be seen, this beam searching in IA will results into a long delay. Moreover, IA occurs more often in mmWave MIMO cellular networks than in conventional systems [5]. Thus, the long IA delay for effective beamforming reduces the feasibility of mmWave MIMO cellular networks.

The IA procedures for conventional systems are not suitable for mmWave MIMO cellular networks. In long term evolution (LTE) systems, omnidirectional transmission is employed for IA and beamforming is employed after IA success [4]. Because the omnidirectional transmission cannot achieve any array gain [3][6][7], IA success can only be achieved for MSs very close to the BS, i.e., this causes a disparity between the range that an MS can be detected and the range that an MS can be served [8]. In mmWave short-range communication systems, e.g., IEEE 802.11ad standard, hierarchical beam codebooks are employed. More specifically, wide-coverage low-gain coarse beams are firstly employed and the best one is selected; Then, narrow-coverage high-gain fine beams for the coverage of the best coarse beam are employed and the best one is selected [3][9]. However, the coarse beams cannot achieve a high array gain and cannot achieve IA success for MSs with long distances to the BS, i.e., this will also cause a disparity between the detectable range and the service range in mmWave cellular networks. Moreover, the collision probability of IA is high with dense user scenario and the IA mechanism is enhanced in [10].

For mmWave MIMO cellular networks, there are three kinds of IA procedures, i.e., the sequential search approaches, the position information (PI) based approaches, and the channel estimation approaches. In sequential search approaches, the basic one is the exhaustive search that sequentially scans all the fine beams and selects the beam with the highest SNR [11]. Suppose the BS array can form N_{BS} fine beams and the MS array can form N_{MS} fine beams, then it will take $N_{\mathrm{BS}}N_{\mathrm{MS}}$ scan slots to select the beam in one side. In order to reduce the IA delay, three kinds of sequential search approaches have been proposed. The first kind is a hierarchical sequential search approach proposed in [12], where the MS firstly uses $N_{\mathrm{MS,C}}$ coarse beams and then uses $N_{\mathrm{MS,F}}$ fine beams to look for the highest SNR. This approach costs $N_{\mathrm{BS}}(N_{\mathrm{MS,C}}+N_{\mathrm{MS,F}})$ scan slots to select the beam in one side. But the IA delay is still too long with large N_{BS}. In [13], it is shown that with the hierarchical codebook in [14] employed at the BS and a posterior probability based beam searching method employed at the MS, the right beamforming vector can be selected with high probability in low signal-to-noise ratio (SNR) scenarios. However, this hierarchical codebook faces the low array gain as those in IEEE 802.11ad standard [3][9]. The second kind is a beam broadening approach proposed in [15], where only

This research was partially supported by Project LY20F010007 supported by Zhejiang Provincial Natural Science Foundation of China, and partially supported by Project 61601152 and Project U1609216 supported by National Natural Science Foundation of China. *(Corresponding author: Anzhong Hu.)*

A. Hu is with the School of Communication Engineering, Hangzhou Dianzi University, Hangzhou 310018, China, email: huaz@hdu.edu.cn.

J. He is with the Centre for Wireless Communications, FI-90014, University of Oulu, Finland, email: jiguang.he@oulu.fi.

the coarse beams are employed and the coarse beams are design to achieve a high array gain for a wide area. The third kind is the semi-directional approach in [16], where the BS or the MS transmits with beamforming and the other is omnidirectional. In [16], the coverage performance with omnidirectional transmission, semi-directional transmission, and fully-directional transmission (i.e., both the BS and the MS are directional) is compared and the results show the superiority of fully-directional transmission.

In order to reduce the IA delay and at the same time employ fully-directional transmission, the PI based approaches have been proposed, where PI refers to the position information of the MS and the BS. In [17], the MS sends its position information to the macro BS with the LTE link, and the macro BS forwards this information to the micro BS with the backhaul link, then the micro BS forms an mmWave beam towards the LOS direction of the MS. Similarly, in [4] and [18], the MS gets the position information of the micro BS and forms a beam towards the LOS direction of the micro BS. By taking into account of the position inaccuracy, a greedy approach that scans around the LOS direction is proposed in [19]. Moreover, the blockage caused by obstacles is addressed in [20], where the beam in an IA success is stored for the MS position. When an MS is close to this position and cannot be detected with the LOS beam, this stored beam will be used for detection. However, with blockage and multipaths, the potential beam direction may differ much from the LOS direction, and the store of the beam for any position necessitates exhaustive search when the use of the LOS beam and the stored beam causes an IA failure. In [21], the authors propose to store the potential beam pairs for each grid area of the serving area and select the potential beam pairs according to the position of the MS. In [22], the beam pairs are learned online with the positions of the MSs in a hierarchy way. However, they did not take into consideration of the MS rotation, and did not analyze the IA delay or probability of miss detection (PMD). In [23], the BS tries one beam in the data base that correspond to the MS position or a few beams that correspond to the nearby positions, but only one beam is stored in the data base for each position, which increases the PMD. The train network is considered in [24], where the authors assume perfect knowledge of the best beam for each position of the MS and predicts the MS position for selecting the beam. Apparently, this assumption is not appropriate for the cellular scenario with fast fading channels. In [25], the authors use machine learning to determine the beam width with the position information and search all the beams. However, it is a kind of exhaustive search and still faces the problem of a long delay or a PMD. In [26], the MS which is a vehicle employs the its position and the light detection and ranging (LIDAR) sensor to select beam pairs for the BS to use. However, the employment of LIDAR requires to deploy the LIDAR sensor on the MS, which increases the cost.

Besides, the channel estimation approaches can also be employed for IA. In [14], a hierarchical beam codebook is proposed and a compressive sensing (CS) based channel estimation approach is proposed. However, each MS should have IA separately, which causes a lot of overhead for the case of a large number of MSs or frequent handovers between cells. To overcome this disadvantage, [27] proposes a CS based channel estimation approach, in which the BS sends time-domain pilots with random beams to all the MSs, each MS estimates the channel, and feedbacks the channel information. But the estimation performance depends on the length of the pilots, which may cause a long delay for a small estimation error.

Towards the goal of overcoming these shortages, in this paper, a PI based beam learning approach is proposed for IA in mmWave MIMO cellular networks. The motivation of employing the learning approach is to utilize the sparse scattering of mmWave to reduce the IA delay. As the mmWave scattering is sparse and is similar for nearby MSs, the beams can be learned for the partitioned areas in the cell. Thus, the beams can be learned efficiently, and the cardinality of the learned beam set can be small. Thus, a short IA delay and a high IA success rate can be achieved. In the proposed approach, the serving area of each micro BS is partitioned into smaller areas according to the corresponding LOS DOAs, and a corresponding empty beam set is formed. In the learning state, the exhaustive search is employed and the success beam will be incorporated into the beam set corresponding to the position of the MS. In the non-learning state, only the learned beam sets are used for the IA. Moreover, the cardinality of each beam set is restricted to prevent long IA delay in the non-learning state. The main contributions of this paper are three-fold.

1) A detailed beam learning procedure is presented. This procedure takes into account of the heterogeneous structure, the MS rotations, the beam selection, and the learning and non-learning periods.
2) An MS classification method based on the sparse scattering of mmWave signal is proposed, which partitions the serving area into smaller areas. For each partitioned area, the beams are learned and incorporated into a potential beam set with a cardinality constraint. Thus, the beams can be learned efficiently and a short IA delay can be achieved.
3) The IA delay is analyzed and compared with other methods. The results show that the proposed IA procedure can achieve a short IA delay. Moreover, the PMD lower bound is derived and the result shows the influence of the system parameters on the PMD lower bound.

This paper is organized as follows. In Section II, the system model and the assumptions are given. Section III presents the proposed IA procedure and the beam learning approach. In Section IV, the IA delay and the PMD are analyzed. Section V gives the simulation parameters and the numerical results. Finally, conclusions are drawn in Section VI.

Notations: Lower-case (upper-case) boldface symbols denote vectors (matrices); $\mathbf{0}_K$ represents the $K \times 1$ zero vector; $(\cdot)^H$ denotes the conjugate transpose; $[\cdot]_j$ is the j-th element of a vector; $|\cdot|$ is the absolute value of a variable or the cardinality of a set; i is the imaginary unit; $\mathrm{P}(\cdot)$ is the probability of an event; $\mathbf{z} \sim \mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ means that $\mathbf{z}$ is a complex circularly symmetric Gaussian vector with mean $\boldsymbol{\mu}$

and covariance matrix $\mathbf{\Sigma}$.

II. System Model

We consider a heterogeneous network which consists of one macro BS and one micro BS as shown in Fig. 1. The macro BS communicates with the micro BS and the MS at LTE frequencies, the micro BS communicates with the MS at mmWave frequencies. There is one MS in this network and this MS is a mobile phone. In fact, the macro BS serves multiple micro BSs and MSs. With the positioning of the MSs and the links operate at LTE frequencies, the macro BS can find the nearest micro BS for each MS and coordinate the IA process for them. In addition, it is possible that the MS and the nearest micro BS is completely blocked so that the IA cannot succeed. Then, the macro BS will arrange the IA between the MS and another nearby micro BS. In this paper, we only consider the IA between one MS and one micro BS that is not completely blocked to the considered MS.

The micro BS is equipped with one uniform rectangular array (URA) and the MS is of one URA. The URAs are working at mmWave frequencies. The URA on the micro BS serves the area of 120° azimuth angular range in front of it. The URA on the micro BS is composed of $N_{\mathrm{BS,H}}$ antenna elements in the horizontal direction and $N_{\mathrm{BS,V}}$ antenna elements in the vertical direction, and these antennas are fully connected to one radio frequency (RF) chain. The URA on the MS is composed of $N_{\mathrm{MS,H}}$ antenna elements in the horizontal direction and $N_{\mathrm{MS,V}}$ antenna elements in the vertical direction, and these antennas are fully connected to one RF chain. Note that the linear array may also be employed at the micro BS or the MS, and the system here can be easily generated to that case. Additionally, when backhaul links are available, the micro BSs can communicate with the macro BS through backhaul links, like the distributed system in [28]. In this case, the LTE links are not necessary.

The channel model in [29] is employed here. The channel at time t is denoted as $\mathbf{H}_t \in \mathbb{C}^{N_{\mathrm{BS,H}}N_{\mathrm{BS,V}}\times N_{\mathrm{MS,H}}N_{\mathrm{MS,V}}}$. Note that t is the index of the time slot, and each time slot is of T_{per} milliseconds (ms). The line-of-sight (LOS) path may exist or not, and the non-LOS (NLOS) paths always exist. When both the LOS path and the NLOS paths exist, the channel from the MS to the micro BS can be denoted as $\mathbf{H}_t = \mathbf{H}_t^{\mathrm{Rician}} \in \mathbb{C}^{N_{\mathrm{BS,H}}N_{\mathrm{BS,V}}\times N_{\mathrm{MS,H}}N_{\mathrm{MS,V}}}$. When only the NLOS paths exist, the channel from the MS to the micro BS can be denoted as $\mathbf{H}_t = \mathbf{H}_t^{\mathrm{Rayleigh}} \in \mathbb{C}^{N_{\mathrm{BS,H}}N_{\mathrm{BS,V}}\times N_{\mathrm{MS,H}}N_{\mathrm{MS,V}}}$. Moreover, we have

$$\mathbf{H}_t^{\mathrm{Ricean}} = \sqrt{\frac{1}{K_{\mathrm{R}}+1}}\beta_t \mathbf{a}_{\mathrm{BS}}(\phi_t,\theta_t)\mathbf{a}_{\mathrm{MS}}^H(\tilde{\phi}_t,\tilde{\theta}_t) + \sqrt{\frac{K_{\mathrm{R}}}{K_{\mathrm{R}}+1}}\sum_{n_{\mathrm{cl}}=1}^{N_{\mathrm{cl}}}\sum_{n_{\mathrm{ray}}=1}^{N_{\mathrm{ray}}}\beta_{n_{\mathrm{cl}},n_{\mathrm{ray}},t} \times \mathbf{a}_{\mathrm{BS}}(\phi_{n_{\mathrm{cl}},n_{\mathrm{ray}},t},\theta_{n_{\mathrm{cl}},n_{\mathrm{ray}},t})\mathbf{a}_{\mathrm{MS}}^H(\tilde{\phi}_{n_{\mathrm{cl}},n_{\mathrm{ray}},t},\tilde{\theta}_{n_{\mathrm{cl}},n_{\mathrm{ray}},t}), \quad (1)$$

where K_{R} is the Ricean K-factor, $\beta_t \sim \mathcal{CN}(0,\sigma_t^2)$ and $\beta_{n_{\mathrm{cl}},n_{\mathrm{ray}},t} \sim \mathcal{CN}(0,\sigma^2_{n_{\mathrm{cl}},n_{\mathrm{ray}},t})$ are the fading coefficients and are independent; ϕ_t and θ_t are the azimuth direction-of-arrival (DOA) and the elevation DOA of the LOS path, $\tilde{\phi}_t$ and $\tilde{\theta}_t$ are the azimuth direction-of-departure (DOD) and the elevation DOD of the LOS path, $\phi_{n_{\mathrm{cl}},n_{\mathrm{ray}},t}$ and $\theta_{n_{\mathrm{cl}},n_{\mathrm{ray}},t}$ are the azimuth DOA and the elevation DOA of the NLOS path of the n_{ray}-th path in the n_{cl}-th cluster, $\tilde{\phi}_{n_{\mathrm{cl}},n_{\mathrm{ray}},t}$ and $\tilde{\theta}_{n_{\mathrm{cl}},n_{\mathrm{ray}},t}$ are the azimuth DOD and the elevation DOD of the NLOS path of the n_{ray}-th path in the n_{cl}-th cluster; N_{cl} and N_{ray} are the number of clusters and the number of paths in each cluster, respectively; $\mathbf{a}_{\mathrm{BS}}(\phi,\theta) \in \mathbb{C}^{N_{\mathrm{BS,H}}N_{\mathrm{BS,V}}\times 1}$ is the micro BS array steering vector, $\mathbf{a}_{\mathrm{MS}}(\phi,\theta) \in \mathbb{C}^{N_{\mathrm{MS,H}}N_{\mathrm{MS,V}}\times 1}$ is the MS array steering vector, and their definition is

$$[\mathbf{a}_{\mathrm{side}}(\phi,\theta)]_{(n-1)N_{\mathrm{S,H}}+m} = \frac{1}{\sqrt{N_{\mathrm{S,H}}N_{\mathrm{S,V}}}}\exp(i\frac{2\pi}{\lambda}d \times((m-1)\sin(\phi)\sin(\theta)+(n-1)\cos(\theta))), \quad (2)$$

where $n = 1,2,\cdots,N_{\mathrm{S,V}}$, $m = 1,2,\cdots,N_{\mathrm{S,H}}$, "side" is either MS or BS, d is the distance between adjacent antennas, and λ is the wavelength. In addition, we have

$$\mathbf{H}_t^{\mathrm{Rayleigh}} = \sum_{n_{\mathrm{cl}}=1}^{N_{\mathrm{cl}}}\sum_{n_{\mathrm{ray}}=1}^{N_{\mathrm{ray}}}\beta_{n_{\mathrm{cl}},n_{\mathrm{ray}},t} \times \mathbf{a}_{\mathrm{BS}}(\phi_{n_{\mathrm{cl}},n_{\mathrm{ray}},t},\theta_{n_{\mathrm{cl}},n_{\mathrm{ray}},t})\mathbf{a}_{\mathrm{MS}}^H(\tilde{\phi}_{n_{\mathrm{cl}},n_{\mathrm{ray}},t},\tilde{\theta}_{n_{\mathrm{cl}},n_{\mathrm{ray}},t}). \quad (3)$$

In addition, for the micro BS $0 \le \theta \le \pi$ and $-\pi/3 \le \phi \le \pi/3$. Moreover, $\theta = \pi/2$ means the direction is perpendicular to the micro BS array, $\theta = 0$ means the direction is right above the micro BS array, and $\theta = \pi$ means the direction is under the micro BS array; $\phi = 0$ means that the direction is the normal of the micro BS array. For clarity, the angles are shown in Fig. 1. In addition, time division duplex is assumed, which means the two-way channels are reciprocal.

Note that the DODs of the MS are defined based on the basis axes on the MS. As the basis axes rotate with the rotation of the MS, the DODs are related to the rotation of the MS. More specifically, when the rotation of the MS changes, i.e., the MS rotates when the time t changes, the DODs change accordingly. Thus, the consideration of the DODs means that the rotation of the MS is considered in this paper. Moreover, it should be noted that rotation tracking is not considered. Thus, the DODs vary over time and are assumed to be unknown. The rotation tracking is left for future work. Also, it is assumed that the position of the MS is available at the MS, which can be obtained with resorting to the positioning system. It is further

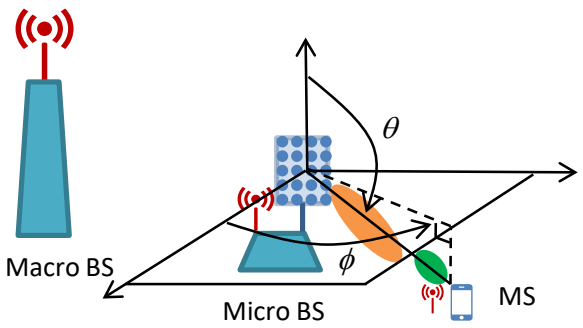

Fig. 1. The illustration of the system and the angles. The red antennas operate at LTE frequencies, the beams are generated by URAs that operate at mmWave frequencies.

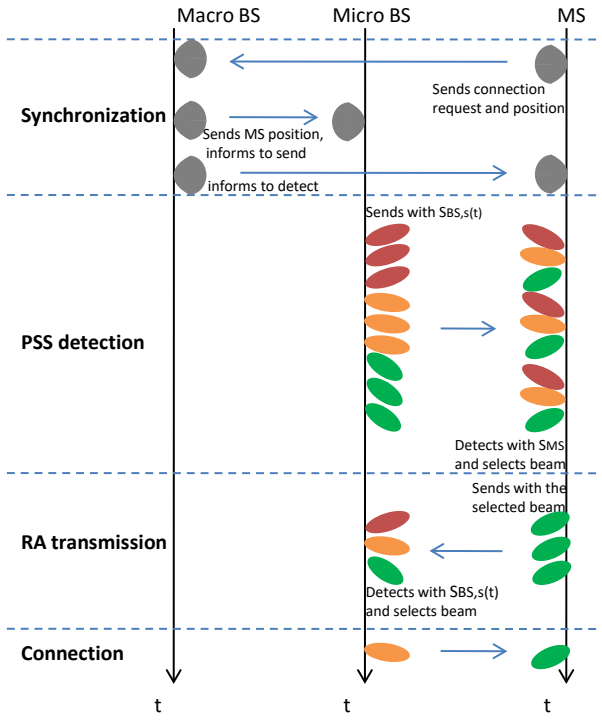

Fig. 2. The illustration of the proposed IA procedure.

assumed that the position information, also termed as PI here, of the MS can be conveyed to the micro BS with the aid of the macro BS at LTE frequencies. Thus, the LOS DOAs, i.e., ϕ_t, θ_t, are known to the micro BS.

Then, in contrast to the conventional beamforming which searches in the potential beam set with the channel state information, which can be obtained with methods such as the channel estimation and the DOA estimation, the position information can be utilized in beam selection in the IA procedure. However, how to choose beams and design the IA procedure to achieve a low PMD and a short IA delay is still a question, and will be investigated in this paper.

III. Beam Learning Based Initial Access Procedure

There are several IA procedures based on PI in [4], [17], [18]. In these approaches, the macro BS controls both the MS and the micro BS in the IA procedure, informs the position information of the MS and the micro BS to each other, and informs the MS or/and the micro BS to steer the beam towards the other one. However, the position error and the blockage of the LOS path deteriorate the beam selection performance. Although these issues are addressed in [19], [20], the exhaustive search is sometimes necessary and results into a long delay.

A. Initial Access Procedure

In contrast to these PI based IA procedures, we propose to inform the position of the MS to the nearest micro BS and the MS should not be in the deaf area of this micro BS [1], and the micro BS searches in a potential beam set. The main difference is that the beam of the micro BS is not simply steered towards the MS but selected from a potential beam set. As long as the potential beam set is properly built for the LOS path and the NLOS paths, the near optimal beam can be selected with a short IA delay. More specifically, the proposed IA procedure is shown in Fig. 2, and is explained as follows.

1) *Synchronization*: The MS sends connection request and its position to the macro BS. The macro BS searches for the micro BS that is closest to this MS and is able to serve this MS, i.e., this MS should not be inside the deaf area of the micro BS. The macro BS sends the position of the MS to the micro BS, informs the micro BS to send primary synchronization signal (PSS), and informs the MS to detect the PSS.
2) *PSS detection*: The micro BS sends PSS once every $T_{\rm per}$ ms. Moreover, the micro BS uses the potential beams in the set $\mathcal{S}_{{\rm BS},s(t)}$, where $\mathcal{S}_{{\rm BS},s(t)}$ is the set that is learned by the micro BS and used for the MS. Note that $s(t)$ stands for the group index for the considered MS and is determined by the position of this MS, and t stands for the time. Meanwhile, the MS uses beams in the beam set $\mathcal{S}_{\rm MS}$ to receive the PSS. As can be seen, this stage will take $|\mathcal{S}_{{\rm BS},s(t)}||\mathcal{S}_{\rm MS}|T_{\rm per}$ ms.
3) *Random access preamble transmission*: The MS transmits a random access (RA) preamble once every $T_{\rm per}$ ms, with the beam that corresponds to the highest SNR in the PSS detection stage. The micro BS sequentially uses beams in the set $\mathcal{S}_{{\rm BS},s(t)}$ to receive the RA preamble. If the highest SNR surpasses the threshold Θ, the corresponding beam is selected for transmission. This stage will take $|\mathcal{S}_{{\rm BS},s(t)}|T_{\rm per}$ ms.
4) *Connection*: If the RA preamble is successfully detected, further connection request and channel scheduling will be conducted. Otherwise, an IA failure occurs and the MS cannot be connected to the micro BS.

B. Beam Learning Problem

From the above comparison, we can see that the proposed PI based IA procedure can achieve a short delay and a low PMD, provided that the potential beam set $\mathcal{S}_{{\rm BS},s(t)}$ is properly constructed. Moreover, a proper beam set should satisfy the following two constraints:

1) *Cardinality*: The cardinality should satisfy $|\mathcal{S}_{{\rm BS},s(t)}| \ll |\mathcal{S}_{\rm BS}|$. More specifically, we assume the beam set of the micro BS is defined as

$$\mathcal{S}_{\rm BS} = \{\mathbf{a}_{\rm BS}(\bar{\phi}_q, \bar{\theta}_{q'}) | \bar{\phi}_q = -\pi/3 + \frac{q-1}{2^{\frac{Q_{\rm B}}{2}}} 2\pi/3, \bar{\theta}_{q'} = \frac{q'-1}{2^{\frac{Q_{\rm B}}{2}}}\pi, q, q' = 1, 2, \cdots, 2^{\frac{Q_{\rm B}}{2}}\}, \quad (4)$$

where $Q_{\rm B}$ is the number of bits that control the phases on the micro BS, and the phase ranges are set according to the serving area of each URA on the micro BS, as shown in

[1]The deaf area of one micro BS refers to a circular region around the micro BS, i.e., the radius is the minimum distance between one MS and the serving micro BS.

Fig. 3. The illustration of the IA and beam learning period.

the second paragraph of this section. As $|\mathcal{S}_{\rm BS}| = 2^{Q_{\rm B}}$, we have $|\mathcal{S}_{{\rm BS},s(t)}| \ll 2^{Q_{\rm B}}$. Moreover, the potential beam set is a subset of the beam set, i.e., $\mathcal{S}_{{\rm BS},s(t)} \subset \mathcal{S}_{\rm BS}$. Here, we restrict the cardinality as $|\mathcal{S}_{{\rm BS},s(t)}| = S, \forall s(t)$.

2) *Beamforming gain*: If the beams that can achieve a high gain are not in $\mathcal{S}_{{\rm BS},s(t)}$, the SNRs in the RA transmission stage will not be high enough for detection, then the PMD will be high. Thus, the beams that may achieve a high gain should be in $\mathcal{S}_{{\rm BS},s(t)}$. Moreover, as the micro BS only has the position information, the beams in $\mathcal{S}_{{\rm BS},s(t)}$ should be learned before the IA procedure.

In this paper, the micro BS conducts beam learning periodically and exploits exhaustive search based approach for IA in the beam learning stage. More specifically, as shown in Fig. 3, in every period of $T_{\rm const}$ ms, the micro BS first conducts exhaustive search based IA for the MS with request, and conducts beam learning at the same time, which takes T_0 ms. Then, the micro BS conducts the proposed IA for the MS with request, which takes $T_{\rm const} - T_0$ ms. In this paper, we assume that the environment changes slowly such that $T_{\rm const} \gg T_0$. Thus, the IA delay is mainly determined by the proposed IA, and we will take the delay of the proposed IA as the delay of the system with the proposed IA. According to the illustrations at the last subsection and the beginning of this subsection, we know that the proposed IA will take $T_{\rm pro} = (|\mathcal{S}_{\rm MS}| + 1)ST_{\rm per}$ ms.

In the beam learning mode, the micro BS transmits PSS with all the beams in $\mathcal{S}_{\rm BS}$, which is the beam set at the micro BS, and the MS receives with all the beams in $\mathcal{S}_{\rm MS}$, which is the beam set at the MS, and is defined as

$$\begin{aligned}\mathcal{S}_{\rm MS} = \{&\mathbf{a}_{\rm MS}(\check{\phi}_p, \check{\theta}_{p'}) | \check{\phi}_p = -\pi/2 + \frac{p-1}{2^{\frac{Q_{\rm M}}{2}}}\pi, \\ &\check{\theta}_{p'} = \frac{p'-1}{2^{\frac{Q_{\rm M}}{2}}}\pi, p, p' = 1, 2, \cdots, 2^{\frac{Q_{\rm M}}{2}}\},\end{aligned} \tag{5}$$

where $Q_{\rm M}$ is the number of bits that control the phases on the MS. Then, the MS transmits RA preamble with the beam in $\mathcal{S}_{\rm MS}$ which corresponds to the highest SNR, and the micro BS receives with all the beams in $\mathcal{S}_{\rm BS}$.With the maximum SNR and the position of the MS, the micro BS learns the potential beam sets for MSs in its serving area.

At time t, the received RA signal at the micro BS can be expressed as

$$\begin{aligned}z_t = &\mathbf{a}_{\rm BS}^H(\phi_{{\rm BS},t}, \theta_{{\rm BS},t})\mathbf{H}_t\mathbf{a}_{\rm MS}(\phi_{{\rm MS},t}, \theta_{{\rm MS},t})s_t \\ &+\mathbf{a}_{\rm BS}^H(\phi_{{\rm BS},t}, \theta_{{\rm BS},t})\mathbf{n}_t,\end{aligned} \tag{6}$$

where s_t is the transmitted symbol, $\mathbf{n}_t \in \mathbb{C}^{N_{\rm BS,H}N_{\rm BS,V}\times 1}$ is the received noise and this vector satisfies $\mathbf{n}_t \sim \mathcal{CN}(\mathbf{0}_{N_{\rm BS,H}N_{\rm BS,V}}, \sigma_{\rm n}^2\mathbf{I}_{N_{\rm BS,H}N_{\rm BS,V}})$; $\mathbf{a}_{\rm BS}(\phi_{{\rm BS},t}, \theta_{{\rm BS},t}) \in \mathcal{S}_{\rm BS}$ is the BS-side beam, $\mathbf{a}_{\rm MS}(\phi_{{\rm MS},t}, \theta_{{\rm MS},t}) \in \mathcal{S}_{\rm MS}$ is the MS-side beam. Moreover, $\mathbf{a}_{\rm MS}(\phi_{{\rm MS},t}, \theta_{{\rm MS},t})$ is selected by choosing the beam corresponding to the maximum SNR when the micro BS transmits to the MS in the PSS stage.

Then, the corresponding SNR at the micro BS is

$$\gamma_t = \frac{|\mathbf{a}_{\rm BS}^H(\phi_{{\rm BS},t}, \theta_{{\rm BS},t})\mathbf{H}_t\mathbf{a}_{\rm MS}(\phi_{{\rm MS},t}, \theta_{{\rm MS},t})s_t|^2}{|\mathbf{a}_{\rm BS}^H(\phi_{{\rm BS},t}, \theta_{{\rm BS},t})\mathbf{n}_t|^2}. \tag{7}$$

Assume that the considered beam learning starts at $t = 0$, and the beam learning period T_0 can be evenly divided into N times of exhaustive search based IA procedure. Then, all the SNRs in the n-th procedure can be written into one set as

$$\begin{aligned}\mathcal{Y}_n = \{&\gamma_t | \mathbf{a}_{\rm BS}(\phi_{{\rm BS},t}, \theta_{{\rm BS},t}) \in \mathcal{S}_{\rm BS}, \mathbf{a}_{\rm MS}(\phi_{{\rm MS},t}, \theta_{{\rm MS},t}) \\ &\in \mathcal{S}_{\rm MS}, (n-1)T_0/T_{\rm per}/N + 1 \le t \le nT_0/T_{\rm per}/N\}.\end{aligned} \tag{8}$$

The problem investigated here is how to design $\mathcal{S}_{{\rm BS},s(t)}$ according to the positions of the MSs and the SNR sets $\mathcal{Y}_n, n = 1, 2, \cdots, N$.

C. MS Classification Based Beam Learning

In this subsection, we will first classify the MS according to its position. Then, we will propose to learn the beams for each group of MSs based on the SNR.

1) MS Classification: When the LOS path exists, the channels corresponding to the LOS paths are highly correlated for MSs with close LOS DOAs. Moreover, the scattering environments are likely to be similar for neighboring MSs, the channels corresponding to the NLOS paths are also highly correlated for MSs with close LOS DOAs. Thus, we can simply classify MSs with close LOS DOAs as one group, e.g., the s-th group, and construct a potential beam set for the MSs in this group, i.e., $\mathcal{S}_{{\rm BS},s(t)}$. Note that the LOS DOA can be calculated by the micro BS with the location information of the MS.

More specifically, by extracting the LOS part in (1), we have

$$\mathbf{H}_t^{\rm LOS} = \sqrt{\frac{1}{K_{\rm R}+1}}\beta_t\mathbf{a}_{\rm BS}(\phi_t, \theta_t)\mathbf{a}_{\rm MS}^H(\tilde{\phi}_t, \tilde{\theta}_t). \tag{9}$$

Substituting this equation into (6), the signal part in the received signal can be denoted as

$$\begin{aligned}y_t &= \mathbf{a}_{\rm BS}^H(\phi_{{\rm BS},t}, \theta_{{\rm BS},t})\mathbf{H}_t^{\rm LOS}\mathbf{a}_{\rm MS}(\phi_{{\rm MS},t}, \theta_{{\rm MS},t}) \\ &= \sqrt{\frac{1}{K_{\rm R}+1}}\beta_t\mathbf{a}_{\rm BS}^H(\phi_{{\rm BS},t}, \theta_{{\rm BS},t})\mathbf{a}_{\rm BS}(\phi_t, \theta_t) \\ &\times\mathbf{a}_{\rm MS}^H(\tilde{\phi}_t, \tilde{\theta}_t)\mathbf{a}_{\rm MS}(\phi_{{\rm MS},t}, \theta_{{\rm MS},t}).\end{aligned}$$

Since the correlation $|\mathbf{a}_{\rm BS}^H(\phi_{{\rm BS},t}, \theta_{{\rm BS},t})\mathbf{a}_{\rm BS}(\phi_t, \theta_t)|$ generally decreases with the increase of $|\phi - \phi_t|$ or $|\theta - \theta_t|$, we can classify MSs according to these angle differences. We define the $s_{q,q'}$-th group as MSs with the LOS DOAs that satisfy

$$\begin{aligned}(\phi_t, \theta_t) \in \mathcal{G}_{s_{q,q'}} = \{&(\phi, \theta) | |\bar{\phi}_q - \phi| \le |\bar{\phi}_{\tilde{q}} - \phi|, \\ &|\bar{\theta}_{q'} - \theta| \le |\bar{\theta}_{\tilde{q}'} - \theta|, \forall \tilde{q}, \tilde{q}'\},\end{aligned} \tag{10}$$

where $s_{q,q'} = 2^{\frac{Q_{\rm B}}{2}}(q-1) + q'$.

2) Beam Learning: Consider the n-th learning procedure, at time t, the SNR of the received signal in (8) is compared with the detection threshold Θ. Additionally, the detection hypotheses are

$$\mathrm{H}_0: \quad \gamma_t \geq \Theta, \quad \exists \gamma_t \in \mathcal{Y}_n, \tag{11}$$

$$\mathrm{H}_1: \quad \gamma_t < \Theta, \quad \forall \gamma_t \in \mathcal{Y}_n, \tag{12}$$

where H_0 represents the signal is present, and H_1 represents the signal is absent. If H_0 holds, the time corresponding to the maximum γ_t is denoted as

$$t_n^\star = \arg\max_t \gamma_t \in \mathcal{Y}_n. \tag{13}$$

Based on the definition, this MS belongs to the $s(t)$-th group according to the LOS DOAs, cf. (10). If the potential beam set is not full, i.e., $|\mathcal{S}_{\mathrm{BS},s(t_n^\star)}| < S$, this beam is added into the potential beam set, and the potential beam set is updated to

$$\mathcal{S}_{\mathrm{BS},s(t_n^\star)} \cup \{\mathbf{a}_{\mathrm{BS}}(\phi_{\mathrm{BS},t_n^\star}, \theta_{\mathrm{BS},t_n^\star})\}. \tag{14}$$

More specifically, the proposed MS classification based beam learning algorithm is presented in Algorithm 1.

Algorithm 1 MS Classification Based Beam Learning

Input: positions of the MSs in the n-th IA procedure for beam learning, $n = 1, 2, \cdots, N$
Output: $\mathcal{S}_{\mathrm{BS},s(t)}, \forall s(t)$
Initialize: $\mathcal{S}_{\mathrm{BS},s(t)} = \emptyset, \forall s(t)$
1: for $n = 1 \to N$
2: calculate the LOS DOAs ϕ_t, θ_t according to the position of the MS
3: classify the MS into each group according to (10)
4: detect with (11) and (12)
5: if H_0 holds and $|\mathcal{S}_{\mathrm{BS},s(t_n^\star)}| < S$
6: calculate the beam vector as (13)
7: update the beam set $\mathcal{S}_{\mathrm{BS},s(t_n^\star)}$ to (14)
8: end
9: end

IV. Performance Analysis and Comparison

A. IA Procedure Comparison

For clarity, we present the brief stages in the other two PI based IA procedures, one exhaustive search based IA procedure, and two iterative search based IA procedures in Table I. Meanwhile, the IA delays, i.e., the times taken by the PSS and RA stages of these procedures, are also presented. In addition, the proposed PI based IA procedure is also shown for comparison. For the proposed PI based IA procedure, if $S \ll |\mathcal{S}_{\mathrm{BS}}|$, the IA delay is much shorter than that of the exhaustive search in [5] and may be comparable to that of the other two PI based approaches. With the simulation parameters in Section V-A, the IA delays of the partial procedures are listed in Table II here. More specifically, the simulation parameters are $S = 4$, $T_{\mathrm{per}} = 5$ ms, $|\mathcal{S}_{\mathrm{MS}}| = 2^{Q_{\mathrm{M}}} = 16$, $|\mathcal{S}_{\mathrm{BS}}| = 2^{Q_{\mathrm{B}}} = 64$. Besides, the delays are not related to the scattering environment such as LOS or NLOS. As can be seen, the IA delay of the proposed PI procedure is comparable to that of other PI based IA procedures and is much shorter than the delay of the exhaustive search procedure. Moreover, if the set $\mathcal{S}_{\mathrm{BS},s(t)}$ is properly constructed, the PMD of the proposed PI based IA procedure can get close to that of the exhaustive approach, and is superior to that of the other two PI based IA approaches when the LOS path does not exist. For the two iterative approaches, their delays depend on the parameters $G, \tilde{S}, \tilde{N}$. For example, with $\tilde{N} = 10, \tilde{S} = 2, G = 7$, the delay of the iterative method in [14] is 1.12 s, and the delay of the iterative method in [13] is 0.8 s. In this case, the delay of the proposed method is shorter than that of the iterative methods. In the case that there is no MS rotation, the LOS DODs are invariant, and the NLOS DODs change more slowly than the case with MS rotation. Then, the MS beam selection is not necessary in each IA procedure, i.e., we can use the same MS beam in several IA procedures. Thus, the effect of MS rotation is that the IA delay should be longer. In the case that the DODs and DOAs of the partial paths are available in the beam learning stage, e.g., the MS is equipped with LIDAR, the micro BS and the MS can use both the position information and the angle information to learn partial beams that correspond to these angles. Thus, the beam learning can be more efficient.

Despite the IA delay, we can also compare the computational complexities. In fact, for all the approaches in Table I, we can see that there is negligible computational complexity. This is because the beam steering is employed with the analog phase shifts, and the sending of the positions or comparing power is of negligible computational complexity. However, for the support vector machine (SVM) based method from [30] and [31], the potential beam set should be found with calculations of the MS position. Since the selection metric should be calculated for all the micro BS beams and with all the learning points. When the number of learning trials is N, the same as the proposed approach, the computational complexity is on the order of $2^{Q_{\mathrm{B}}} N$.

B. PMD Analysis

In this subsection, we first analyze the PMD in the single path case, and then analyze the asymptotic PMD in the multipath case.

At time t, the received RA signal at the micro BS can be expressed as (6), the corresponding SNR at the micro BS is (7). Here, we analyze the PMD in the RA transmission stage of the proposed IA procedure, in which $\mathbf{a}_{\mathrm{BS}}(\phi_{\mathrm{BS},t}, \theta_{\mathrm{BS},t}) \in \mathcal{S}_{\mathrm{BS},s(t)}$.

Lemma 1. *Assume there is only one path in the channel with*

$$\mathbf{H}_t = \beta_t \mathbf{a}_{\mathrm{BS}}(\phi_t, \theta_t) \mathbf{a}_{\mathrm{MS}}^H(\tilde{\phi}_t, \tilde{\theta}_t), \tag{15}$$

where $\beta_t \sim \mathcal{CN}(0, \sigma_t^2)$, and the proposed IA starts at time T_1, the PMD of the proposed IA satisfies

$$\begin{aligned} &\mathrm{P}(\gamma_t \leq \Theta, \forall T_1 \leq t \leq T_1 + |\mathcal{S}_{\mathrm{MS}}|S - 1) \\ &\geq \prod_{t=T_1}^{T_1+|\mathcal{S}_{\mathrm{MS}}|S-1} \int_0^\infty h(y)(1 - e^{-u_t y/2})\mathrm{d}y, \end{aligned} \tag{16}$$

where $u_t = \Theta\sigma_{\mathrm{n}}^2/(\sigma_t^2 |\tilde{r}_t s_t|^2)$, $\tilde{r}_t = r(f(\tilde{\phi}_t, \tilde{\theta}_t, \phi_{\mathrm{MS},t}, \theta_{\mathrm{MS},t}), N_{\mathrm{MS,H}}) r(g(\tilde{\theta}_t, \theta_{\mathrm{MS},t}), N_{\mathrm{MS,V}})$, $r(x, N) = |\sin(xN/2)/\sin(x/2)|$, $h(x) = 1/2e^{-x/2}$.

TABLE I
BRIEF STAGES AND DELAYS OF IA PROCEDURES

IA procedure	Brief stages	Delay
PI [4] [a]	• The macro BS spreads the positions of the micro BSs to the MS. • The MS selects the closest micro BS and steers towards the LOS direction. All micro BSs search exhaustively in $\mathcal{S}_{\mathrm{BS}}$.	$\lvert\mathcal{S}_{\mathrm{BS}}\rvert T_{\mathrm{per}}$
PI [19] [a,b]	• The MS sends its position to the macro BS, which further assigns the closest micro BS for connection and forwards the MS position. • The micro BS steers towards the LOS direction of the MS. The MS searches exhaustively in $\mathcal{S}_{\mathrm{MS}}$.	$\lvert\mathcal{S}_{\mathrm{MS}}\rvert T_{\mathrm{per}}$
Exhaustive search [5]	• The micro BSs send signal using beams in $\mathcal{S}_{\mathrm{BS}}$. The MS receives with beams in $\mathcal{S}_{\mathrm{MS}}$. • The MS sends RA preamble with the selected beam. The micro BS searches exhaustively in $\mathcal{S}_{\mathrm{BS}}$.	$(\lvert\mathcal{S}_{\mathrm{MS}}\rvert+1)$ $\times\lvert\mathcal{S}_{\mathrm{BS}}\rvert T_{\mathrm{per}}$
Iterative [14]	•for $g=1\rightarrow G$ • The micro BS sends signal using beams in $\mathcal{S}(g,k_g)$ with $\lvert\mathcal{S}(g,k_g)\rvert=\tilde{S}$. The MS receives with beams in $\mathcal{S}_{\mathrm{MS}}$. • The MS sends the received signal to the micro BS with the macro BS. The micro BS calculates k_{g+1}. • end	$G\tilde{S}\lvert\mathcal{S}_{\mathrm{MS}}\rvert T_{\mathrm{per}}$
Iterative [13]	•for $\tilde{n}=1\rightarrow\tilde{N}$ • The micro BS sends signal using the beam corresponding to $D_{\tilde{n}}$. The MS receives with beams in $\mathcal{S}_{\mathrm{MS}}$. • The MS sends the received signal with the maximal power to the micro BS with the macro BS. The micro BS calculates $D_{\tilde{n}+1}$. • end	$\tilde{N}\lvert\mathcal{S}_{\mathrm{MS}}\rvert T_{\mathrm{per}}$
Proposed PI [c]	• The MS sends its position to the macro BS, which further assigns the closest micro BS for connection and forwards the MS position. • The micro BS sends signal using beams in $\mathcal{S}_{\mathrm{BS},s(t)}$. The MS receives with beams in $\mathcal{S}_{\mathrm{MS}}$. • The MS sends RA preamble with the selected beam. The micro BS receives with beams in $\mathcal{S}_{\mathrm{BS},s(t)}$.	$(\lvert\mathcal{S}_{\mathrm{MS}}\rvert+1)$ $\times ST_{\mathrm{per}}$

[a] Because of the finite choices of analog beamforming vector, the beam for the LOS direction is the one that is mostly correlated to the array steering vector corresponding to the LOS path.
[b] The procedure therein is modified in two aspects. First, the MS should scan all possible directions. Second, the beam width at the micro BS is fixed.
[c] In case $\lvert\mathcal{S}_{\mathrm{BS},s(t)}\rvert<S$ after beam leaning, the selected beam in $\mathcal{S}_{\mathrm{BS},s(t)}$ is repeated in the IA to make $\lvert\mathcal{S}_{\mathrm{BS},s(t)}\rvert=S$.

TABLE II
TYPICAL IA DELAYS

IA	PI [4]	PI [19]	Exhaustive search [5]	Proposed PI
Delay (s)	0.32	0.08	5.44	0.34

Proof. Refer to Appendix A. □

As can be seen from Lemma 1, with the increase of the signal power or the decrease of the detection threshold, the PMD decreases. Then, we will analyze the PMD in the multipath case.

Lemma 2. *Denote the existence of the LOS path as $l_t=1$ and the absence of the LOS path as $l_t=0$. Then, the PMD is lower bounded as*

$$\begin{aligned}&\mathrm{P}(\gamma_t\leq\Theta,\forall T_1\leq t\leq T_1+\lvert\mathcal{S}_{\mathrm{MS}}\rvert S-1)\\&\geq\prod_{t=T_1}^{T_1+\lvert\mathcal{S}_{\mathrm{MS}}\rvert S-1}(\int_0^{\infty}h(y)(1-e^{-w_t y/2})\mathrm{d}y l_t\\&+\int_0^{\infty}h(y)(1-e^{-v_t y/2})\mathrm{d}y(1-l_t)),\end{aligned}\tag{17}$$

where $v_t=\Theta\sigma_{\mathrm{n}}^2/(\sum_{n_{\mathrm{cl}}=1}^{N_{\mathrm{cl}}}\sum_{n_{\mathrm{ray}}=1}^{N_{\mathrm{ray}}}\sigma_{n_{\mathrm{cl}},n_{\mathrm{ray}},t}^2\lvert\tilde{r}_{n_{\mathrm{cl}},n_{\mathrm{ray}},t}\rvert^2\lvert s_t\rvert^2)$,

$$\begin{aligned}w_t=\Theta\sigma_{\mathrm{n}}^2/(K_{\mathrm{R}}/(K_{\mathrm{R}}+1)\sum_{n_{\mathrm{cl}}=1}^{N_{\mathrm{cl}}}\sum_{n_{\mathrm{ray}}=1}^{N_{\mathrm{ray}}}\sigma_{n_{\mathrm{cl}},n_{\mathrm{ray}},t}^2\\\times\lvert\tilde{r}_{n_{\mathrm{cl}},n_{\mathrm{ray}},t}\rvert^2\lvert s_t\rvert^2+1/(K_{\mathrm{R}}+1)\sigma_t^2\lvert\tilde{r}_t\rvert^2\lvert s_t\rvert^2).\end{aligned}$$

Proof. Refer to Appendix B. □

As can be seen from Lemma 2, with the increase of the signal power or the decrease of the detection threshold, the PMD lower bound decreases. Moreover, when the number of clusters N_{cl} increases, the PMD lower bound will also decrease. This means that rich scattering is helpful for IA.

V. NUMERICAL RESULTS

A. Simulation Parameters

The parameters in the urban micro-cell (street canyon) scenario of [29] are employed for the system parameters and are shown in Table III. The LOS probability, i.e., the probability of (1), is denoted as

$$Pr_{\mathrm{LOS},t}=\begin{cases}1, & d_{\mathrm{2D},t}\leq 10\text{ m},\\ \frac{18}{d_{\mathrm{2D},t}}+\exp\left(-\frac{d_{\mathrm{2D},t}}{36}\right) & \\ \times(1-\frac{18}{d_{\mathrm{2D},t}}), & 10\text{ m}<d_{\mathrm{2D},t},\end{cases}\tag{18}$$

where $d_{\mathrm{2D},t}$ is the horizontal distance between the MS and the micro BS. Thus, the NLOS probability, i.e., the probability of (3), is $1-Pr_{\mathrm{LOS},t}$. The micro BS antenna height is denoted as h_{BS}. The MS height is denoted as h_t and is uniformly distributed in the range $[1.5, 22.5]$ m. The MS distributes uniformly in the horizontal annulus of inner radius $r_{\min}$ and outer radius $r_{\max}$. Additionally, the micro BS is in the center of the annulus. The carrier frequency is denoted as f_{c}. The micro BS transmission power and the MS transmission power are both denoted as $P_{\mathrm{T}}=s_t^2$, and the noise power is denoted as $P_{\mathrm{n}}=\sigma_{\mathrm{n}}^2$.

The antennas on the micro BS and the MS are omnidirectional. For the LOS path, σ_t below (1) accounts the path loss PL_t and the shadow fading SF_t as

$$10\log_{10}(\sigma_t)=-PL_t-SF_t,\tag{19}$$

TABLE III
SIMULATION PARAMETERS

Parameter	Value	Parameter	Value
d/λ	0.5	N	100
h_{BS}	10 m	K_{R}	$10^{0.9}$
T_{per}	5 ms	f_{c}	28 GHz
σ_{SF}	4 dB	σ'_{SF}	8.2 dB
N_{cl}	5	N_{ray}	20
$N_{\mathrm{BS,H}}$	8	$N_{\mathrm{BS,V}}$	8
$N_{\mathrm{MS,H}}$	4	$N_{\mathrm{MS,V}}$	4
Θ	-4 dB	Q_{B}	6
Q_{M}	4	S	4
c_{ASA}	22°	P_{T}	23 dBm
P_{n}	-174 dBm	$r_{\min}$	10 m
$r_{\max}$	100 m		

where

$$PL_t = \begin{cases} 32.4 + 21\log_{10}(d_t) + 20\log_{10}(f_{\mathrm{c}}), \\ \qquad 10\text{ m} \le d_{\mathrm{2D},t} \le D_t, \\ 32.4 + 40\log_{10}(d_t) + 20\log_{10}(f_{\mathrm{c}}) \\ -9.5\log_{10}(D_t^2 + (h_{\mathrm{BS}} - h_t)^2), \\ \qquad D_t \le d_{\mathrm{2D},t} \le 5\text{ km}, \end{cases}$$

d_t is the distance between the MS and the micro BS, $D_t = 4(h_{\mathrm{BS}}-1)(h_t-1)f_{\mathrm{c}}/c$ is the breakpoint distance, $c = 3.0\times10^8$ m/s. For the shadow fading, SF_t is Gaussian distributed with zero mean and variance σ^2_{SF}.

For the NLOS path, $\sigma_{n_{\mathrm{cl}},n_{\mathrm{ray}},t}$ below (1) accounts the path loss $PL_{n_{\mathrm{cl}},t}$ and the shadow fading $SF_{n_{\mathrm{cl}},t}$ as

$$10\log_{10}(\sigma_{n_{\mathrm{cl}},n_{\mathrm{ray}},t}) = -PL'_t - SF'_{n_{\mathrm{cl}},t}, \tag{20}$$

where

$$PL'_t = 32.4 + 31.9\log_{10}(d_t) + 20\log_{10}(f_{\mathrm{c}}), \tag{21}$$

$SF'_{n_{\mathrm{cl}},t}$ is Gaussian distributed with zero mean and variance ${\sigma'_{\mathrm{SF}}}^2$. The azimuth DOAs of the NLOS paths are generated as

$$\phi_{n_{\mathrm{cl}},n_{\mathrm{ray}},t} = \phi_{n_{\mathrm{cl}},t} + c_{\mathrm{ASA}}\alpha_{n_{\mathrm{cl}},n_{\mathrm{ray}},t}, \tag{22}$$

where $\phi_{n_{\mathrm{cl}},t}$ is the mean of the azimuth DOA in the n_{cl}-th cluster; c_{ASA} is the cluster-wise RMS DOA spread, $\alpha_{n_{\mathrm{cl}},n_{\mathrm{ray}},t}$ is the ray offset DOA defined in Table 7.5-3 of [29]. Moreover, the multipath DOAs are in a pattern as follows. Denote the position of the MS at time t in the planer axes as x, y. For $x \le (r_{\min}+r_{\max})/2$ and $y \ge 0$, $-60° \le \phi_{n_{\mathrm{cl}},t} \le -56°$; For $x \le (r_{\min}+r_{\max})/2$ and $y < 0$, $-30° \le \phi_{n_{\mathrm{cl}},t} \le -26°$; For $x > (r_{\min}+r_{\max})/2$ and $y \ge 0$, $26° \le \phi_{n_{\mathrm{cl}},t} \le 30°$; For $x > (r_{\min}+r_{\max})/2$ and $y < 0$, $56° \le \phi_{n_{\mathrm{cl}},t} \le 60°$. Additionally, the elevation DOAs, the azimuth DODs, and the elevation DODs of the NLOS paths are generated in the same way as that for the azimuth DOAs. In Section V-B, we first simulate the extreme case that the LOS probability is 1 and there is no multipath. This corresponds to the scenario that there is no reflecting surfaces and the MS is in the visible region of the micro BS. In Section V-C, we simulate the normal case that the LOS appears randomly with a probability in (18) and there are multipaths. This corresponds to the scenario that there are random blockages and many reflecting surfaces.

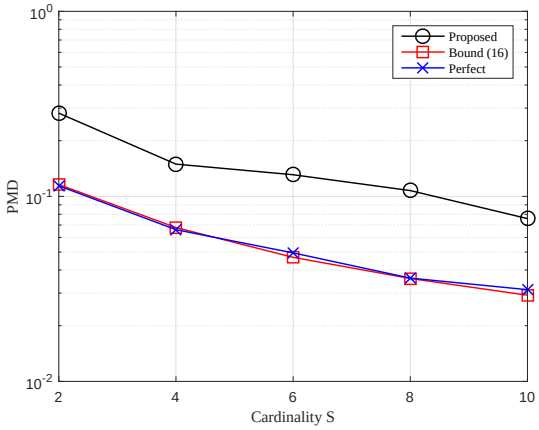

Fig. 4. The PMD versus the potential beam cardinality S.

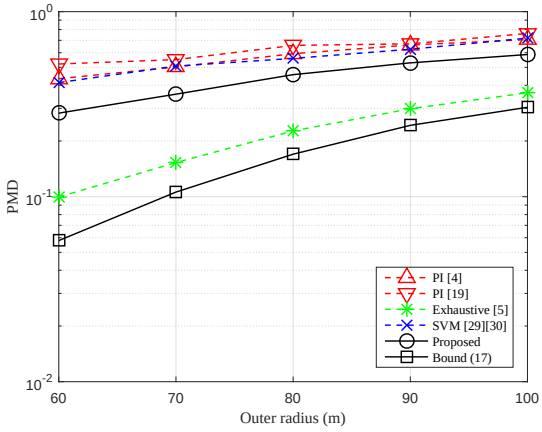

Fig. 5. The PMD versus the outer distance between the MS and the micro BS $r_{\max}$.

B. Results With Single LOS Path

In this subsection, we set the LOS probability as $Pr_{\mathrm{LOS},t} = 1$ and the number of multipaths as $N_{\mathrm{cl}} = 0$. We simulate the relation between the PMD with the potential beam cardinality S. Moreover, the PMD lower bound in (16) is also simulated. The results are shown in Fig. 4. In this figure, the PMD decreases with the increase of S. This is because the BS array can steer the beam direction in a more precise way with the increase of S, and correspondingly the array gain achieved is higher. The PMD of the proposed method is close to the lower bound in (16). This result shows that the derived lower bound is effective. Also, the PMD with the maximal array gain is also simulated, which is denoted as "Perfect". The PMD with the maximal array gain is the almost the same as the lower bound, which also shows the effectiveness of the derived lower bound.

C. Results With Multipaths

In this subsection, we compare the PMDs of various methods in a multipath scenario. The compared methods include the PI based method in [4], the PI based method in [19], the exhaustive search method in [5], as shown in Table I.

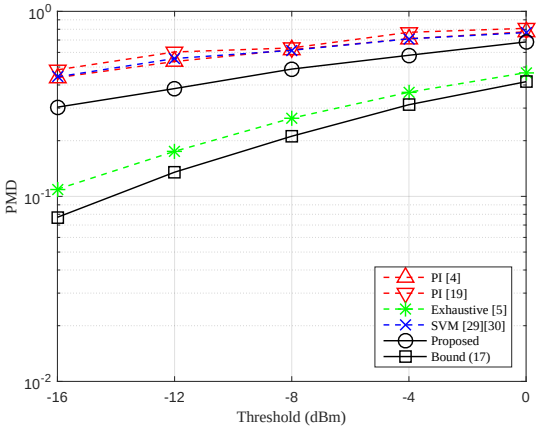

Fig. 6. The PMD versus the detection threshold Θ.

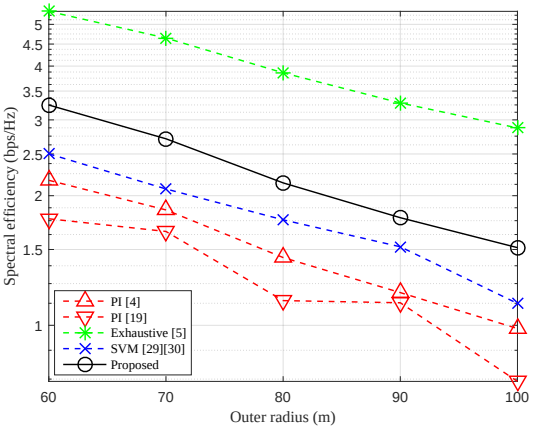

Fig. 8. The spectral efficiency versus the outer distance between the MS and the micro BS $r_{\max}$.

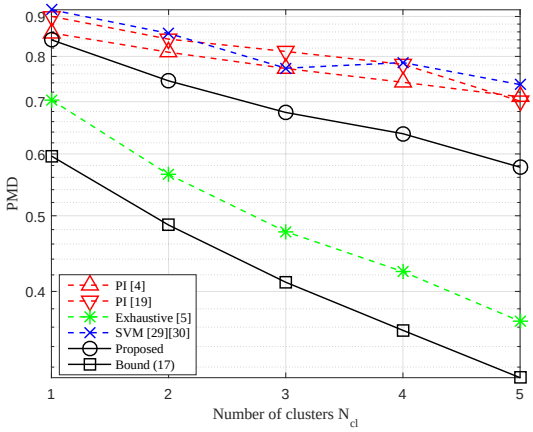

Fig. 7. The PMD versus the number of clusters.

Moreover, we also simulated the SVM based method from [30] and [31]. Note that the slots with exhaustive searches for beam learning in the proposed PI method and the SVM based method are the same. Moreover, the cardinality of the potential beam set used in the proposed PI method is the same as that for the maximal number of beams in the SVM method. Thus, we keep fairness between the proposed PI method and the SVM method. Note that the SVM method is also based on the position information available, is thus a PI based method. Moreover, the PMD lower bound in (17) is also simulated.

In Fig. 5, the outer radius of the cell area increases, and all the PMDs increase at the same time. The reason is that the distances between the MSs and the micro BS increase and the propagation losses increase simultaneously. The proposed PI method performs better than other methods except the exhaustive search method in the range $[60\ 100]$ m. This result shows that the proposed method can achieve a higher PMD than other PI based methods. Meanwhile, the proposed PI method performs similar as other methods except the exhaustive search method when the outer radius is 100 m. This is because most of the MSs are not in the serving area of the micro BS and almost all the beams are not effective, which can be verified by the small gap between these PI methods and the exhaustive search method.

In Fig. 6, the detection threshold increases, and all the PMDs increase at the same time. The reason is that a higher detection threshold requires a higher SNR for an IA success. The proposed PI method performs better than other PI methods, especially in the low detection threshold regime. Meanwhile, in all the simulations, the PMD lower bound is lower than the PMDs and changes in the same rate as the PMDs of these methods. These results verify the effectiveness of the derived lower bound.

In Fig. 7, the PMD versus the number of clusters is shown. As can be seen, the PMDs decrease with the increase of the number of clusters. This is because the probability that the beam can achieve a high gain increases with the number of clusters. Meanwhile, it can be seen that the PMD of the proposed method is lower than other PI based methods.

In the above results, we can see that the PMD bound is not quite close to the PMD of the proposed method. However, the PMD bound is the same as that of the "Perfect" beam in the single path scenario and is close to that of the exhaustive search. The reason is as follows. We use the maximum micro BS array gain to derive the PMD bound, the proposed method cannot always select the best beam, and the best beam which is selected by the exhaustive cannot achieve the maximum micro BS array gain for the limit value of the BS beam cardinality $|\mathcal{S}_{\mathrm{BS}}|$. Additionally, the "Perfect" beam is in fact the case with maximum micro BS array gain achieved.

In Fig. 8, we compare the spectral efficiency when the outer radius of the cell area increases. Note the spectral efficiency is calculated with the SNR in (7), and it is only calculated when the detection succeed. It can be seen that the spectral efficiency decreases with the increase of the outer distance. This is because the path loss increases when the outer distance increases. Moreover, it can be seen that the spectral efficiency of the proposed approach is higher than that of other approaches except the exhaustive search. This result also shows the superiority of the proposed approach.

VI. Conclusions

In this paper, the beam selection problem in the IA state of mmWave MIMO cellular networks is investigated. Based on the sparse scattering property of mmWave communication, the serving area of one BS is partitioned into smaller areas. By exploiting the position information of the MS and the selected beams in the learning stage, the potential beams are learned. In the non-learning stage, the learning results are used for beam selection. The analysis shows the lower bound of the proposed method, and the simulation shows that the proposed beam selection method can achieve a lower PMD than other PI based methods. Meanwhile, the IA delay of the proposed method is comparable to other PI based methods and is more practical for implementation than the exhaustive search method.

Appendix A
Proof of Lemma 1

According to Fig. 3, the proposed method will use the exhaustive search and the proposed IA which uses the learned beams. Following the analysis in section III-B, the received signal expression and the SNR expression of the proposed IA are the same as that of the exhaustive search in (6) and (7). The difference is that in the exhaustive search, the BS beam satisfies $\mathbf{a}_{\mathrm{BS}}(\phi_{\mathrm{BS},t},\theta_{\mathrm{BS},t})\in\mathcal{S}_{\mathrm{BS}}$, while in the proposed IA, the beam satisfies $\mathbf{a}_{\mathrm{BS}}(\phi_{\mathrm{BS},t},\theta_{\mathrm{BS},t})\in\mathcal{S}_{\mathrm{BS},s(t)}$. Assume that the proposed IA starts at time T_1, the PMD of the proposed IA is

$$\begin{aligned}&\mathrm{P}(\gamma_t\le\Theta,\forall T_1\le t\le T_1+|\mathcal{S}_{\mathrm{MS}}|S-1)\\&=\prod_{t=T_1}^{T_1+|\mathcal{S}_{\mathrm{MS}}|S-1}\mathrm{P}(\gamma_t\le\Theta).\end{aligned}\tag{23}$$

Substituting (15) into (6) yields

$$\begin{aligned}z_t&=\beta_t\mathbf{a}_{\mathrm{BS}}^H(\phi_{\mathrm{BS},t},\theta_{\mathrm{BS},t})\mathbf{a}_{\mathrm{BS}}(\phi_t,\theta_t)\mathbf{a}_{\mathrm{MS}}^H(\tilde{\phi}_t,\tilde{\theta}_t)\\&\times\mathbf{a}_{\mathrm{MS}}(\phi_{\mathrm{MS},t},\theta_{\mathrm{MS},t})s_t+\mathbf{a}_{\mathrm{BS}}^H(\phi_{\mathrm{BS},t},\theta_{\mathrm{BS},t})\mathbf{n}_t\\&=\beta_t\frac{1}{N_{\mathrm{BS,H}}N_{\mathrm{BS,V}}}\frac{1-\exp(iN_{\mathrm{BS,H}}f(\phi_t,\theta_t,\phi_{\mathrm{BS},t},\theta_{\mathrm{BS},t}))}{1-\exp(if(\phi_t,\theta_t,\phi_{\mathrm{BS},t},\theta_{\mathrm{BS},t}))}\\&\times\frac{1-\exp(iN_{\mathrm{BS,V}}g(\theta_t,\theta_{\mathrm{BS},t}))}{1-\exp(ig(\theta_t,\theta_{\mathrm{BS},t}))}\\&\times\frac{1}{N_{\mathrm{MS,H}}N_{\mathrm{MS,V}}}\frac{1-\exp(iN_{\mathrm{MS,H}}f(\tilde{\phi}_t,\tilde{\theta}_t,\phi_{\mathrm{MS},t},\theta_{\mathrm{MS},t}))}{1-\exp(if(\tilde{\phi}_t,\tilde{\theta}_t,\phi_{\mathrm{MS},t},\theta_{\mathrm{MS},t}))}\\&\times\frac{1-\exp(iN_{\mathrm{MS,V}}g(\tilde{\theta}_t,\theta_{\mathrm{MS},t}))}{1-\exp(ig(\tilde{\theta}_t,\theta_{\mathrm{MS},t}))}s_t+\mathbf{a}_{\mathrm{BS}}^H(\phi_{\mathrm{BS},t},\theta_{\mathrm{BS},t})\mathbf{n}_t,\end{aligned}$$

where

$$\begin{aligned}f(\phi_t,\theta_t,\phi_{\mathrm{BS},t},\theta_{\mathrm{BS},t})&=\frac{2\pi}{\lambda}d(\sin(\phi_t)\sin(\theta_t)\\&\quad-\sin(\phi_{\mathrm{BS},t})\sin(\theta_{\mathrm{BS},t})),\\g(\theta_t,\theta_{\mathrm{BS},t})&=\frac{2\pi}{\lambda}d(\cos(\theta_t)-\cos(\theta_{\mathrm{BS},t})).\end{aligned}$$

By applying

$$\frac{1}{N}\left|\frac{1-\exp(ixN)}{1-\exp(ix)}\right|=\frac{1}{N}\left|\frac{\sin(xN/2)}{\sin(x/2)}\right|\triangleq r(x,N)\tag{24}$$

to the above equation, we have

$$\begin{aligned}z_t&=\beta_t r(g(\theta_t,\theta_{\mathrm{BS},t}),N_{\mathrm{BS,V}})\\&\times r(f(\phi_t,\theta_t,\phi_{\mathrm{BS},t},\theta_{\mathrm{BS},t}),N_{\mathrm{BS,H}})\\&\times r(f(\tilde{\phi}_t,\tilde{\theta}_t,\phi_{\mathrm{MS},t},\theta_{\mathrm{MS},t}),N_{\mathrm{MS,H}})\\&\times r(g(\tilde{\theta}_t,\theta_{\mathrm{MS},t}),N_{\mathrm{MS,V}})s_t+\mathbf{a}_{\mathrm{BS}}^H(\phi_{\mathrm{BS},t},\theta_{\mathrm{BS},t})\mathbf{n}_t,\end{aligned}$$

and the corresponding SNR as

$$\gamma_t=\frac{|z_t-\mathbf{a}_{\mathrm{BS}}^H(\phi_{\mathrm{BS},t},\theta_{\mathrm{BS},t})\mathbf{n}_t|^2}{|\mathbf{a}_{\mathrm{BS}}^H(\phi_{\mathrm{BS},t},\theta_{\mathrm{BS},t})\mathbf{n}_t|^2}.\tag{25}$$

Since $r(x,N)\le 1$, we have the upper bound of the SNR as

$$\gamma_t\le\gamma_t^{\mathrm{b}}=\frac{|\beta_t\tilde{r}_t s_t|^2}{|\mathbf{a}_{\mathrm{BS}}^H(\phi_{\mathrm{BS},t},\theta_{\mathrm{BS},t})\mathbf{n}_t|^2},\tag{26}$$

where

$$\tilde{r}_t=r(f(\tilde{\phi}_t,\tilde{\theta}_t,\phi_{\mathrm{MS},t},\theta_{\mathrm{MS},t}),N_{\mathrm{MS,H}})r(g(\tilde{\theta}_t,\theta_{\mathrm{MS},t}),N_{\mathrm{MS,V}}).$$

Then, we have

$$\begin{aligned}&\mathrm{P}(\gamma_t\le\Theta)\ge\mathrm{P}(\gamma_t^{\mathrm{b}}\le\Theta)\\&=\mathrm{P}(\frac{2|\beta_t|^2}{\sigma_t^2}-\frac{2\Theta|\mathbf{a}_{\mathrm{BS}}^H(\phi_{\mathrm{BS},t},\theta_{\mathrm{BS},t})\mathbf{n}_t|^2}{\sigma_t^2|\tilde{r}_t s_t|^2}\le 0).\end{aligned}$$

Since $\beta_t\sim\mathcal{CN}(0,\sigma_t^2)$ and $\mathbf{n}_t\sim\mathcal{CN}(\mathbf{0}_{N_{\mathrm{BS,H}}N_{\mathrm{BS,V}}},\sigma_{\mathrm{n}}^2\mathbf{I}_{N_{\mathrm{BS,H}}N_{\mathrm{BS,V}}})$, $2|\beta_t|^2/\sigma_t^2$ is Chi-square distributed with two degrees of freedom, $2|\mathbf{a}_{\mathrm{BS}}^H(\bar{\phi}_q,\bar{\theta}_{q'})\mathbf{n}_t|^2/\sigma_{\mathrm{n}}^2$ is Chi-square distributed with two degrees of freedom. Thus, we have

$$\begin{aligned}\mathrm{P}(\gamma_t^{\mathrm{b}}\le\Theta)&=\int_0^\infty h(y)\int_0^{u_t y}h(x)\mathrm{d}x\mathrm{d}y\\&=\int_0^\infty h(y)(1-e^{-u_t y/2})\mathrm{d}y,\end{aligned}$$

where $u_t=\Theta\sigma_{\mathrm{n}}^2/(\sigma_t^2|\tilde{r}_t s_t|^2)$, $h(x)=1/2e^{-x/2}$ is the probability distribution function of a Chi-square distributed variable with two degrees of freedom. Thus, we have (16).

Appendix B
Proof of Lemma 2

Since $r(x,N)\le 1$, with the NLOS multipath channel model in (3), we have

$$\begin{aligned}&|\mathbf{a}_{\mathrm{BS}}^H(\phi_{\mathrm{BS},t},\theta_{\mathrm{BS},t})\mathbf{H}_t\mathbf{a}_{\mathrm{MS}}(\phi_{\mathrm{MS},t},\theta_{\mathrm{MS},t})|\\&\le\sum_{n_{\mathrm{cl}}=1}^{N_{\mathrm{cl}}}\sum_{n_{\mathrm{ray}}=1}^{N_{\mathrm{ray}}}\beta_{n_{\mathrm{cl}},n_{\mathrm{ray}},t}\tilde{r}_{n_{\mathrm{cl}},n_{\mathrm{ray}},t},\end{aligned}$$

where

$$\begin{aligned}\tilde{r}_{n_{\mathrm{cl}},n_{\mathrm{ray}},t}&=r(g(\tilde{\theta}_{n_{\mathrm{cl}},n_{\mathrm{ray}},t},\theta_{\mathrm{MS},t}),N_{\mathrm{MS,V}})\\&\times r(f(\tilde{\phi}_{n_{\mathrm{cl}},n_{\mathrm{ray}},t},\tilde{\theta}_{n_{\mathrm{cl}},n_{\mathrm{ray}},t},\phi_{\mathrm{MS},t},\theta_{\mathrm{MS},t}),N_{\mathrm{MS,H}}).\end{aligned}$$

Correspondingly, we have

$$\gamma_t\le\frac{|\sum_{n_{\mathrm{cl}}=1}^{N_{\mathrm{cl}}}\sum_{n_{\mathrm{ray}}=1}^{N_{\mathrm{ray}}}\beta_{n_{\mathrm{cl}},n_{\mathrm{ray}},t}\tilde{r}_{n_{\mathrm{cl}},n_{\mathrm{ray}},t}s_t|^2}{|\mathbf{a}_{\mathrm{BS}}^H(\bar{\phi}_q,\bar{\theta}_{q'})\mathbf{n}_t|^2}.$$

Because $\beta_{n_{\rm cl},n_{\rm ray},t} \sim \mathcal{CN}(0,\sigma^2_{n_{\rm cl},n_{\rm ray},t})$ and is independent of each other, we have

$$\sum_{n_{\rm cl}=1}^{N_{\rm cl}}\sum_{n_{\rm ray}=1}^{N_{\rm ray}} \beta_{n_{\rm cl},n_{\rm ray},t}\tilde{r}_{n_{\rm cl},n_{\rm ray},t} \sim \mathcal{CN}(0,\sum_{n_{\rm cl}=1}^{N_{\rm cl}}\sum_{n_{\rm ray}=1}^{N_{\rm ray}} \sigma^2_{n_{\rm cl},n_{\rm ray},t}|\tilde{r}_{n_{\rm cl},n_{\rm ray},t}|^2).$$

Then, similar to the derivations in the proof of Lemma 1, we have

$$\begin{aligned}&\mathrm{P}(\gamma_t \le \Theta, \forall T_1 \le t \le T_1 + |\mathcal{S}_{\rm MS}|S-1)\\ \ge &\prod_{t=T_1}^{T_1+|\mathcal{S}_{\rm MS}|S-1}\int_0^\infty h(y)(1-e^{-v_t y/2})\mathrm{d}y,\end{aligned}$$

where

$$v_t = \Theta\sigma_{\rm n}^2/(\sum_{n_{\rm cl}=1}^{N_{\rm cl}}\sum_{n_{\rm ray}=1}^{N_{\rm ray}} \sigma^2_{n_{\rm cl},n_{\rm ray},t}|\tilde{r}_{n_{\rm cl},n_{\rm ray},t}|^2|s_t|^2).$$

Similarly, for the case with both the LOS and NLOS paths, the PMD satisfies

$$\begin{aligned}&\mathrm{P}(\gamma_t \le \Theta, \forall T_1 \le t \le T_1 + |\mathcal{S}_{\rm MS}|S-1)\\ \ge &\prod_{t=T_1}^{T_1+|\mathcal{S}_{\rm MS}|S-1}\int_0^\infty h(y)(1-e^{-w_t y/2})\mathrm{d}y,\end{aligned}$$

where

$$\begin{aligned}w_t = \Theta\sigma_{\rm n}^2/(K_{\rm R}/(K_{\rm R}+1)\sum_{n_{\rm cl}=1}^{N_{\rm cl}}\sum_{n_{\rm ray}=1}^{N_{\rm ray}}\sigma^2_{n_{\rm cl},n_{\rm ray},t}\\ \times|\tilde{r}_{n_{\rm cl},n_{\rm ray},t}|^2|s_t|^2 + 1/(K_{\rm R}+1)\sigma_t^2|\tilde{r}_t|^2|s_t|^2).\end{aligned}$$

Then, the lower bound of the PMD is (17).

REFERENCES

[1] T. S. Rappaport, E. Ben-Dor, J. N. Murdock, and Y. Qiao, "38 GHz and 60 GHz angle-dependent propagation for cellular & peer-to-peer wireless communications," in *Proc. 2012 IEEE Int. Conf. Commun.*, Ottawa, Canada, 2012, pp. 4568–4573.

[2] Z. Pi and F. Khan, "An introduction to millimeter-wave mobile broadband systems," *IEEE Commun. Mag.*, vol. 49, no. 6, Jun. 2011, pp. 101–107.

[3] 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, Oct. 2017.

[4] M. Giordani, M. Mezzavilla, and M. Zorzi, "Initial access in 5G mmWave cellular networks," *IEEE Commun. Mag.*, vol. 54, no. 11, pp. 40–47, Nov. 2016.

[5] C. N. Barati, S. A. Hosseini, M. Mezzavilla, T. Korakis, S. S. Panwar, S. Rangan, and M. Zorzi, "Initial access in millimeter wave cellular systems," *IEEE Trans. Wireless Commun.*, vol. 15, no. 12, pp. 7926–7940, Dec. 2016.

[6] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. K. Soong, and J. C. Zhang, "What will 5G be?" *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, Jun. 2014.

[7] 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, Jun. 2014.

[8] C. N. Barati, S. A. Hosseini, S. Rangan, P. Liu, T. Korakis, S. S. Panwar, and T. S. Rappaport, "Directional cell discovery in millimeter wave cellular networks," *IEEE Trans. Wireless Commun.*, vol. 14, no. 12, pp. 6664–6678, Dec. 2015.

[9] P. Zhou, K. Cheng, X. Han, X. Fang, Y. Fang, R. He, Y. Long, and Y. Liu, "IEEE 802.11ay-based mmWave WLANs: Design challenges and solutions," *IEEE Commun. Surveys & Tutorials*, vol. 20, no. 3, pp. 1654–1681, Mar. 2018.

[10] P. Zhou, X. Fang, Y. Fang, Y. Long, R. He, and X. Han, "Enhanced random access and beam training for millimeter wave wireless local networks with high user density," *IEEE Trans. Wireless Commun.*, vol. 16, no. 12, pp. 7760–7773, Dec. 2017.

[11] C. Jeong, J. Park, and H. Yu, "Random access in millimeter-wave beamforming cellular networks: Issues and approaches," *IEEE Commun. Mag.*, vol. 53, no. 1, pp. 180–185, Jan. 2015.

[12] V. Desai, L. Krzymien, P. Sartori, W. Xiao, A. Soong, and A. Alkhateeb, "Initial beamforming for mmwave communications," in *Proc. 48th Asilomar Conf. Signals, Systems and Computers*, Pacific Grove, USA, 2014, pp. 1926–1930.

[13] S. Chiu, N. Ronquillo, and T. Javidi, "Active learning and CSI acquisition for mmWave initial alignment," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 11, pp. 2474–2489, Nov. 2019

[14] A. Alkhateeb, O. E. Ayach, G. Leus, and R. W. Heath, Jr., "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 831–846, Oct. 2014.

[15] V. Raghavan, J. Cezanne, S. Subramanian, A. Sampath, and O. Koymen, "Beamforming tradeoffs for initial UE discovery in millimeter-wave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 543–559, Apr. 2016.

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

[17] Q. C. Li, H. Niu, G. Wu, and R. Q. Hu, "Anchor-booster based heterogeneous networks with mmWave capable booster cells," in *Proc. 2013 IEEE Globecom Workshops*, Atlanta, USA, 2013, pp. 93–98.

[18] W. B. Abbas and M. Zorzi, "Context information based initial cell search for millimeter wave 5G cellular networks," in *Proc. 25th Euro. Conf. Networks and Commun.*, Athens, Greece, 2016, pp. 111–116.

[19] A. Capone, I. Filippini, and V. Sciancalepore, "Context information for fast cell discovery in mmWave 5G networks," in *Proc. 21th Euro. Wireless Conf.*, Budapest, Hungary, 2015, pp. 1–6.

[20] A. Capone, I. Filippini, V. Sciancalepore, and D. Tremolada, "Obstacle avoidance cell discovery using mm-waves directive antennas in 5G networks," in *Proc. IEEE 26th Annual Int. Symp. Personal, Indoor, and Mobile Radio Commun.*, Hong Kong, China, 2015, pp. 2349–2353.

[21] V. Va, J. Choi, T. Shimizu, G. Bansal, and R. W. Heath, Jr., "Inverse multipath fingerprinting for millimeter wave V2I beam alignment," *IEEE Trans. Veh. Technol.*, vol. 67, no. 5, pp. 4042–4058, May 2018.

[22] V. Va, T. Shimizu, G. Bansal, and R. W. Heath, Jr., "Online learning for position-aided millimeter wave beam training," *IEEE Access*, vol. 7, pp. 30507–30526, Mar. 2019.

[23] J. C. Aviles and A. Kouki, "Position-aided mm-wave beam training under NLOS conditions," *IEEE Access*, vol. 4, pp. 8703–8714, Nov. 2016.

[24] J. Kim and A. F. Molisch, "Enabling gigabit services for IEEE 802.11ad-capable high-speed train networks," in *Proc. 2013 IEEE Radio and Wireless Symp.*, Austin, USA, Jan. 2013, pp. 145–147.

[25] R. Zia-ul-Mustafa and S. A. Hassan, "Machine learning-based context aware sequential initial access in 5G mmWave systems," in *Proc. 2019 IEEE Globecom Workshops*, Waikoloa, USA, Dec. 2019, pp. 1–6.

[26] A. Klautau, N. González-Prelcic, and R. W. Heath, Jr., "LIDAR data for deep learning-based mmWave beam-selection," *IEEE Wireless Commun. Lett.*, vol. 8, no. 3, pp. 909–912, Jun. 2019.

[27] X. Song, S. Haghighatshoar, and G. Caire, "Efficient beam alignment for millimeter wave single-carrier systems with hybrid MIMO transceivers," *IEEE Trans. Wireless Commun.*, vol. 18, no. 3, pp. 1518–1533, Mar. 2019.

[28] X. Zhang, D. Guo, K. An, and B. Zhang, "Secure communications over cell-free massive MIMO networks with hardware impairments," *IEEE Syst. J.*, vol. 14, no. 2, pp. 1909–1920, Jun. 2020.

[29] 3GPP, "Study on channel model for frequencies from 0.5 to 100 GHz," 3rd Generation Partnership Project, Tech. Rep. 38.901 V14.3.0, Dec. 2017.

[30] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learning*, vol. 20, no. 3, pp. 273–297, 1995.

[31] Y. Long, Z. Chen, J. Fang, and C. Tellambura, "Data-driven-based analog beam selection for hybrid beamforming under mm-Wave channels," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 2, pp. 340–352, May 2018.

Anzhong Hu (S'13-M'17) received the B.Eng. in communication engineering from Zhejiang University of Technology (ZJUT), Hangzhou, China, in 2009, and the Ph.D. in signal and information processing from Beijing University of Posts and Telecommunications (BUPT), Beijing, China, in 2014. He joined Hangzhou Dianzi University (HDU), Hangzhou, China, in 2014, where he is currently an associate professor. From 2019 to 2020, he was a visiting researcher with Chalmers University of Technology (CTH), Gothenburg, Sweden. His research expertise include signal processing, system design, and parameter optimization in massive multiple-input multiple output (MIMO) systems and millimeter wave (mmWave) systems.

Jiguang He (S'16, M'20) received the B.Eng. degree from Harbin Instituteof Technology, Harbin, China, in 2010, M.Sc. degree from Xiamen University, Xiamen, China, in 2013, and Ph.D. degree from University of Oulu, Oulu, Finland, in 2018, all in communications engineering. From September 2013 to March 2015, he was with State Key Laboratory of Terahertz and Millimeter Waves at City Universityof Hong Kong, working on beam tracking over millimeter wave MIMO systems. Since June 2015, he has been with Centre for Wireless Communications (CWC), University of Oulu, participating in multiple national and international projects, e.g., EU FP7 RESCUE, EU H2020 ARIADNE, AoF Flagship 6Genesis, etc. His research interests span millimeter wave MIMO communications, reconfigurable intelligent surfaces for joint communication and positioning, cooperative relaying, position awareness for communications, etc.