Efficient Beam Training and Channel Estimation for Millimeter Wave Communications Under Mobility

Sun Hong Lim, Jisu Bae, Sunwoo Kim, Byonghyo Shim*, and Jun Won Choi
Hanyang University, *Seoul National University
Email: shlim, jsbae@spo.hanyang.ac.kr, remero@hanyang.ac.kr,
*bshim@snu.ac.kr, junwchoi@hanyang.ac.kr

Abstract

In this paper, we propose an efficient beam training technique for millimeter-wave (mmWave) communications. In the presence of the mobile users under high mobility, the conventional beam training should be performed more frequently to allow the users to acquire channel state information (CSI) accurately. Since it demands high resource overhead for beam training, we introduce the *dedicated beam training* protocol which sends the training beams separately to a specific high mobility user (called a target user) without changing the periodicity of the conventional beam training. The dedicated beam training does not require much resource since only a small number of the training beams are sent to the target user. In order to achieve good system performance with low training overhead, we design the optimal beam selection strategy which finds the best beamforming vectors yielding the lowest channel estimation error based on the target user's probabilistic channel information. Such dedicated beam training is combined with the new greedy channel estimator which effectively estimates the mmWave channel accounting for sparse characteristics and dynamics of the target user's channel. Our numerical evaluation demonstrates that the proposed beam training scheme can maintain good channel estimation performance with significantly less training overhead than the conventional beam training protocols.

Index Terms

Millimeter wave communications, Beam training, Beam selection, Beamforming, Angle of departure (AoD), Mobility

Efficient Beam Training and Channel Estimation for Millimeter Wave Communications Under Mobility

I. Introduction

In recent years, wireless communications using millimeter-wave (mmWave) frequency band has received great deal of attention as a means to meet ever-increasing throughput demand of next generation communication systems [1]–[3]. Basically, millimeter-wave band covers the frequency band ranging from 30 GHz to 300 GHz, which is much higher than the frequency band in current cellular systems. As shown in Fig. 1, there is a vast amount of bandwidth that has not been explored by the current communication systems. Notwithstanding the great promise and potential benefit, there are some drawbacks and obstacles that need to be addressed for the commercialization of mmWave-based communication systems. One major obstacle is the significant path loss of mmWave channels [4]–[6]. Compared to conventional communication systems using microwave radio waves, mmWave band experiences high atmospheric attenuation when the transmit signal is absorbed by gas and humidity. Additionally, there would be a significant path loss when the signal is blocked by obstacles such as building, foliage, and user's body [2]. The key enabler to overcome this drawback is the *beamforming* technique in which two communication entities transmit and receive the signals with appropriately adjusted phase and amplitude using an array of antennas [3], [7]–[9]. Since the wavelength in the mmWave communication systems is in the range of one to ten millimeters, a large number of antenna elements can be integrated into a small form factor, enabling the highly directional beamforming to compensate for the large path loss of mmWave channels.

Over the years, various beamforming strategies, such as hybrid beamforming, switched beamforming, and multi-stage beamforming, have been introduced [10]–[13]. In a nutshell, conventional beamforming protocol consists of two main steps. The first step is the *beam training* [3], [9]. In this step, the basestation transmits the known training symbols to the mobile users periodically. Using the received beams, a user acquires the channel state information (CSI) which

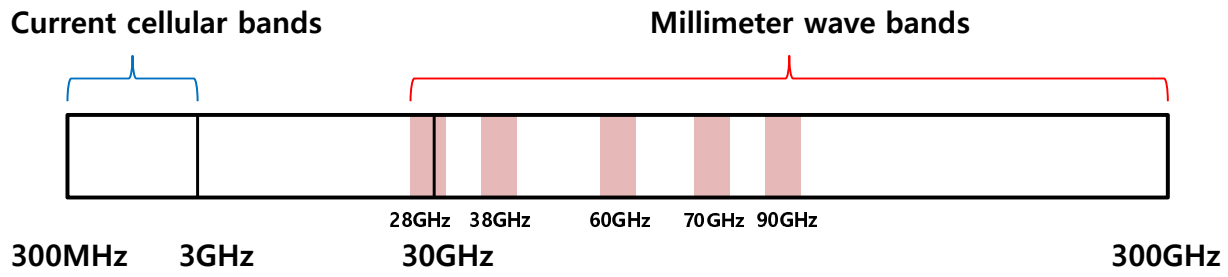


Fig. 1. Unutilized frequency spectrum for wireless communications.

corresponds to the channel gains for all pairs of the transmit antenna and receive antenna. In the widely used *beam-cycling* scheme, for example, the basestation sequentially transmits the N beams steered at the equally distributed directions. In the mobile terminal, each user estimates its own CSI and then feeds back the estimated CSI to the basestation. Using the obtained CSIs, the basestation performs the precoding of data symbols, which corresponds to the second stage of beamforming. In this stage, the basestation transmits the precoded symbols using the precoding matrix designed to maximize the system throughput [14], [15].

While high directional beamforming is an effective means to improve the system performance, the system becomes inefficient when the users' locations change in time or the position and heading angle of the devices vary relative to the base-station. This is because the beam training of the mobile users should be performed more frequently to track the CSIs of the mobile users. Since the beams are shared by all co-scheduled users, even when only a fraction of users are under mobility, more frequent beam training should be performed, thus increasing the training overhead. Recently, various attempts have been made to enhance the beam training efficiency. These include the adaptive beamtraining [9], codebook-based beam switching [16], simplex optimization-based beam training [17], beamcoding approach [18], and multi-level beam training [19]. Also, various beam tracking algorithms to estimate the time-varying mmWave channel have been proposed. These include the probabilistic beam tracking [23], adaptive beamformer and combiner [20], Kalman-based beam tracking [21], and beamspace-based beam tracking [22].

The primary purpose of this paper is to propose a new beam training strategy to support mobility scenario in mmWave communications. The key idea of the proposed scheme is to use the *dedicated beam training* to facilitate the beam training targeted for high mobility user

without changing the periodicity of the conventional beam training. While all users in the cell are supported by the conventional beam referred to as the *common beam*, the proposed dedicated beam is designed to support the users under high mobility that cannot acquire the CSI accurately only using the common beam training. Since the dedicated beam training is intended for the moving user, we can improve the quality of beam training by exploiting the information on the moving user's channel dynamics. As a result, even with a small number of training beams, good system performance for the target user can be ensured. In this perspective, the proposed dedicated beam transmission is analogous to the user-specific pilot transmission scheme in the 4G long-term evolution (LTE) standard [24].

Our contributions in this paper are as follows:

- We propose a new type of beam for mmWave systems called dedicated beam. The proposed dedicated beam training protocol achieves a significant reduction in training overhead by transmitting only a small number of training beams over the radio resource allocated for high mobility users. Note that the dedicated beam training process is activated only when there are high mobility users in a cell.
- We propose a sparse channel estimation technique suited for the dedicated beam training. The proposed technique is based on modified greedy sparse recovery algorithm which successively finds the estimate of each pair of the angle of departure (AoD) and the angle of arrival (AoA) until it finds those corresponding to multi-paths. Based on the received training beams, the proposed channel estimator produces the probabilistic distribution of AoA and AoD, which are used to perform the beam selection and channel estimation for the next beam training period.
- Using the statistical channel model that captures the dynamic channel behavior, we derive the optimal beam selection strategy searching for the best beam indices minimizing the channel estimation error from the beam codebook. Specifically, we derive a computationally efficient beam search strategy for dual beam transmission. We demonstrate from numerical experiments that the combination of our channel estimation algorithm and the dedicated beam training yields the performance comparable to the conventional beam cycling with much smaller training overhead.

It is also worth mentioning that our approach is distinct from the existing beam tracking

techniques in [21], [30]. These methods consider the transmission of a single beam towards the direction of AoD, which is estimated using the AoD tracking algorithm such as the extended Kalman filter. We observe that this single beam transmission strategy does not perform well for high mobility scenario since an incorrect AoD estimation causes a degradation in the received SNR and hence the AoD estimation quality. Such vicious circle could lead to gradual degradation of the channel estimation until when the beam cycling (involving multiple beam transmissions) is performed. We also find that the sparse channel estimation based on the quantized AoD model [3] does not work with the single beam transmission due to poor conditioning of system matrix. On the contrary, our beam training substantially mitigates this problem by using optimally placed multiple beams (at least two) for beam transmission and adopting the enhanced sparse channel estimation accounting for the probabilistic distribution of the dynamic channel.

The rest of this paper is organized as follows; In section II, we introduce the system and channel models for mmWave communications. In section III, we describe the proposed beam tracking method. In section IV, we present the simulation results, and conclude the paper in section V.

The notations to be used in the rest of paper are as follows. Operations $(\cdot)^T$, $(\cdot)^H$, $(\cdot)^*$, and $(\cdot)^{-1}$ denote transpose, Hermitian, conjugate, and inverse operations, respectively. $\mathbf{A} \otimes \mathbf{B}$ is Kronecker product of two matrices $\mathbf{A}$ and $\mathbf{B}$. $E[X]$ and $Var(X)$ denotes the expectation and variance of the random variable X. $(\mathbf{X})_{i,j}$ denotes the (i,j)th element of the matrix $\mathbf{X}$ and $(\mathbf{x})_i$ is the ith element of the vector $\mathbf{x}$. $\mathbf{X}_\Omega$ is the submatrix of $\mathbf{X}$ that contains the columns as specified in the set Ω. and $(\mathbf{x})_\Omega$ is the vector constructed by picking the elements from the vector $\mathbf{x}$ as specified in the set Ω. $A \succeq B$ implies that $(A-B)$ is positive semidefinite. $\mathrm{vec}(\mathbf{X})$ is the vectorization operation of the matrix $\mathbf{X}$. $\mathcal{CN}(\mu, \Sigma)$ denotes the complex multivariate Gaussian distribution

$$f(\mathbf{x}) = \frac{1}{|\pi\Sigma|} \exp\left(-(\mathbf{x}-\mu)^H \Sigma^{-1} (\mathbf{x}-\mu)\right). \tag{1}$$

II. mmWave Channel Model and Conventional Beam Training

In this section, we briefly describe the system model for mmWave communications. We first present the mmWave channel model and then discuss the basic procedure of the conventional beam training and channel estimation.

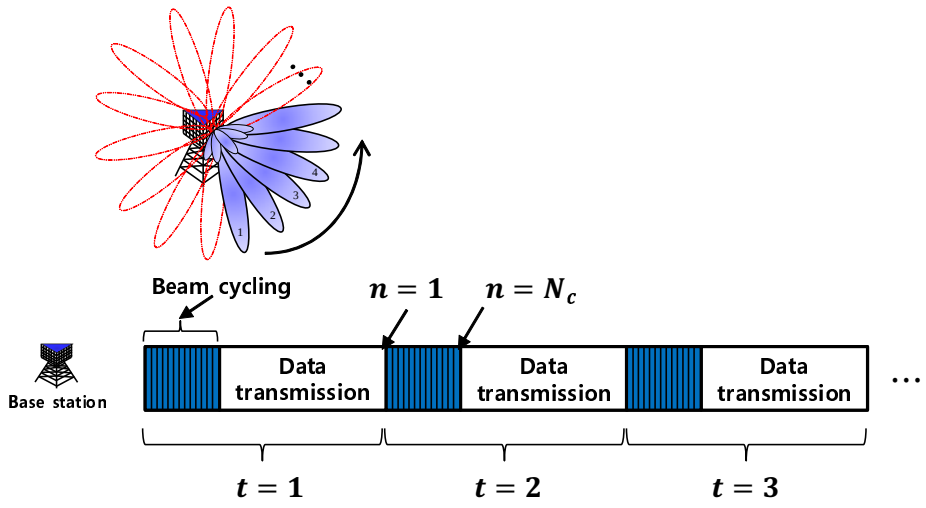


Fig. 2. Frame structure of the conventional common beam training.

A. System Description

Consider the mmWave downlink where a basestation equipped with N_b antennas is serving P users with N_m antennas. In each beam training period, the basestation transmits N_c beams sequentially at equally distributed directions. Each user estimates its own channel based on the received training signals and then feeds back the channel information to the basestation. After generating the precoding matrix using the users' channel state, the basestation transmits the precoded data symbols. This beam training procedure is performed periodically to keep track of channels for all users. The frame structure for the conventional common beam training is depicted in Fig. 2.

B. mmWave Channel Model

In general, the mmWave channel from the basestation to the pth user can be expressed as the $N_m \times N_b$ channel matrix $\mathbf{H}_{p,t,n}$ where the subscripts t and n represent the tth beam training period and the nth beam transmission, respectively (see Fig. 2). The (i, j)th element of $\mathbf{H}_{p,t,n}$ represents the channel gain from the jth antenna of the basestation to the ith antenna of the

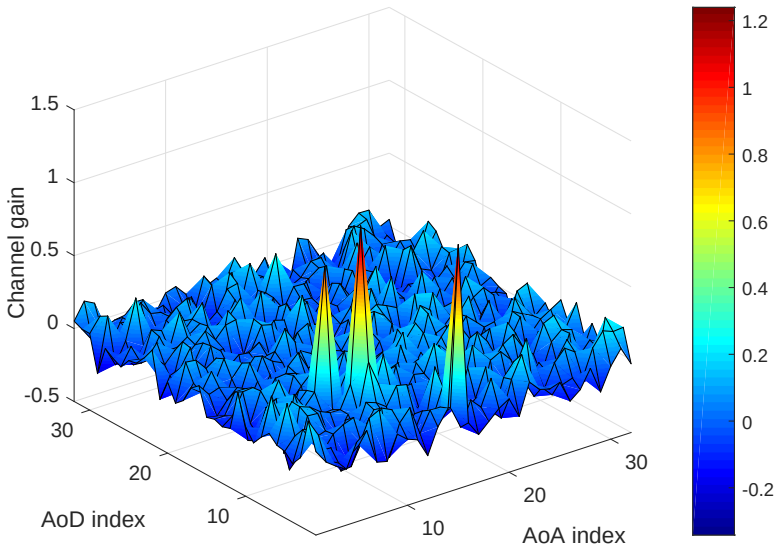


Fig. 3. The distribution of angular domain channel.

user. The angular domain representation of the channel matrix $\mathbf{H}_{p,t,n}$ is expressed as [9]

$$\mathbf{H}_{p,t,n} = \sum_{l=1}^{L} \alpha_{p,t,n,l} \mathbf{a}^{(m)}(\theta^{(m)}_{p,t,n,l}) \left(\mathbf{a}^{(b)}(\theta^{(b)}_{p,t,n,l})\right)^H \tag{2}$$

$$= \mathbf{A}^{(m)}_{p,t,n} \mathbf{\Lambda}_{p,t,n} (\mathbf{A}^{(b)}_{p,t,n})^H, \tag{3}$$

where

$$\mathbf{A}^{(m)}_{p,t,n} = \left[\mathbf{a}^{(m)}(\theta^{(m)}_{p,t,n,1}) \quad \cdots \quad \mathbf{a}^{(m)}(\theta^{(m)}_{p,t,n,L})\right] \tag{4}$$

$$\mathbf{\Lambda}_{p,t,n} = \begin{bmatrix} \alpha_{p,t,n,1} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \alpha_{p,t,n,L} \end{bmatrix} \tag{5}$$

$$\mathbf{A}^{(b)}_{p,t,n} = \left[\mathbf{a}^{(b)}(\theta^{(b)}_{p,t,n,1}) \quad \cdots \quad \mathbf{a}^{(b)}(\theta^{(b)}_{p,t,n,L})\right]. \tag{6}$$

Note that L is the total number of the paths, $\alpha_{p,t,n,l}$ is the channel gain for the lth path, and $\theta^{(b)}_{p,t,n,l}$ and $\theta^{(m)}_{p,t,n,l}$ are the AoD and AoA for the lth path where

$$\theta^{(b)}_{p,t,n,l} = \sin(\phi^{(b)}_{p,t,n,l}),$$
$$\theta^{(m)}_{p,t,n,l} = \sin(\phi^{(m)}_{p,t,n,l})$$

where $\phi^{(b)}_{p,t,n,l}$, $\phi^{(m)}_{p,t,n,l} \in [-\frac{\pi}{2}, \frac{\pi}{2}]$ are the angles in radian for AoD and AoA, respectively. Notice that $\mathbf{a}^{(b)}(\theta)$ and $\mathbf{a}^{(m)}(\theta)$ are the steering vectors for the basestation and user, respectively. That is,

$$\mathbf{a}^{(b)}(\theta) = \frac{1}{\sqrt{N_b}}\left[1, e^{\frac{j2\pi d_b\theta}{\lambda}}, e^{\frac{j2\pi 2d_b\theta}{\lambda}}, \cdots, e^{\frac{j2\pi(N_b-1)d_b\theta}{\lambda}}\right]^T$$
$$\mathbf{a}^{(m)}(\theta) = \frac{1}{\sqrt{N_m}}\left[1, e^{\frac{j2\pi d_m\theta}{\lambda}}, e^{\frac{j2\pi 2d_m\theta}{\lambda}}, \cdots, e^{\frac{j2\pi(N_m-1)d_m\theta}{\lambda}}\right]^T,$$

where d_b and d_m are the distances between the adjacent antennas for the basestation and user, respectively, and λ is the signal wavelength. It is worth mentioning that although the AoD and AoA have real values in the range $[-1, 1]$, they can be approximated to the discrete values by quantizing them on the uniform grid of M_b and M_m bins. That is,

$$\theta^{(b)}_{p,t,n,l} \in \left[-1, -1+2\frac{1}{M_b}, \ldots, -1+2\frac{(M_b-1)}{M_b}\right]$$
$$\theta^{(m)}_{p,t,n,l} \in \left[-1, -1+2\frac{1}{M_m}, \ldots, -1+2\frac{(M_b-1)}{M_m}\right]. \tag{7}$$

Adopting the quantized channel representation, we have so-called virtual channel representation [9]

$$\mathbf{H}_{p,t,n} = \mathbf{A}^{(m)}\mathbf{H}^{(v)}_{p,t,n}\left(\mathbf{A}^{(b)}\right)^H, \tag{8}$$

where

$$\mathbf{A}^{(b)} = \left[\mathbf{a}^{(b)}(-1), \mathbf{a}^{(b)}\Big(-1+\frac{2}{M_b}\Big), \ldots, \mathbf{a}^{(b)}\Big(-1+2\frac{M_b-1}{M_b}\Big)\right]$$
$$\mathbf{A}^{(m)} = \left[\mathbf{a}^{(m)}(-1), \mathbf{a}^{(m)}\Big(-1+\frac{2}{M_m}\Big), \ldots, \mathbf{a}^{(m)}\Big(-1+2\frac{M_m-1}{M_m}\Big)\right].$$

The (i,j)th element of $\mathbf{H}^{(v)}_{p,t,n}$ is the channel gain corresponding to the ith angular bin for the AoA and the jth angular bin for the AoD. If there exist L multi-paths, only L entries of $\mathbf{H}^{(v)}_{p,t,n}$ have dominant values and the rest entries are close to zero. Since the number of multi-paths L is in general much smaller than that of elements in $\mathbf{H}^{(v)}_{p,t,n}$, as illustrated in Fig. 3, the channel matrix $\mathbf{H}^{(v)}_{p,t,n}$ can be readily modeled as a sparse matrix in the angular domain.

C. Conventional Beam Training and Channel Estimation

As mentioned, during the beam training period, the basestation transmits the known symbols over the N_c consecutive time steps. In the nth beam transmission, for example, the basestation transmits the known symbol s_n through the beamforming vector $\mathbf{f}_{t,n}$. The corresponding received vector for the pth user is given by

$$\mathbf{r}_{p,t,n} = \mathbf{H}_{p,t,n}\mathbf{f}_{t,n}s_n + \mathbf{n}_{p,t,n}, \tag{9}$$

where $\mathbf{f}_{t,n}$ is the $N_b \times 1$ beamforming vector and $\mathbf{n}_{p,t,n}$ is the $N_m \times 1$ Gaussian noise vector $\mathcal{CN}(0, \sigma_n^2\mathbf{I})$. Since s_n is the known symbol, we let $s_n = 1$ in the sequel for simplicity. In the generation of the beamforming vector, the beam-cycling scheme in which the basestation transmits N beams at equally spaced directions is popularly used (see Fig.2)

$$\mathbf{f}_{t,1} = \mathbf{a}^{(b)}(-1) \tag{10}$$

$$\mathbf{f}_{t,2} = \mathbf{a}^{(b)}\left(-1 + 2\frac{1}{N_c}\right) \tag{11}$$

$$\vdots$$

$$\mathbf{f}_{t,N_c} = \mathbf{a}^{(b)}\left(-1 + 2\frac{N_c - 1}{N_c}\right) \tag{12}$$

Clearly, the larger the number of beams N_c, the better the quality of the channel estimation would be. However, large N_c leads to higher resource overhead and consequently lower data throughput. The pth user multiplies the $N_m \times 1$ combining vectors $\mathbf{W}_{p,t} = [\mathbf{w}_{p,t,1}, ..., \mathbf{w}_{p,t,N_m}]$ to $\mathbf{r}_{p,t,n}$, i.e.,

$$\mathbf{y}_{p,t,n} = \mathbf{W}_{p,t}^H\mathbf{r}_{p,t,n} \tag{13}$$

$$= \mathbf{W}_{p,t}^H\mathbf{H}_{p,t,n}\mathbf{f}_{t,n}s_n + \mathbf{W}_{p,t}^H\mathbf{n}_{p,t,n}, \tag{14}$$

where $\mathbf{W}_{p,t}$ is called the combining matrix. Well-known example of the combining matrix is the $N_m \times N_m$ DFT matrix $\mathbf{W}_{p,t}$ given by

$$\mathbf{W}_{p,t} = \left[\mathbf{a}^{(m)}(-1), \ldots, \mathbf{a}^{(m)}\left(-1 + 2\frac{N_m - 1}{N_m}\right)\right]. \tag{15}$$

Since the combining matrix is multiplied to the single received vector, increasing the number of the combining vectors (i.e., the column vectors) does not increase the resource overhead in contrast to the beamforming vectors.

Next, we describe the basic channel estimation algorithm used in the conventional beam training. Using the vectors $\mathbf{y}_{p,t,1}, ..., \mathbf{y}_{p,t,N_c}$ collected from the tth beam training period, the pth user estimates the channel matrix $\mathbf{H}_{p,t,n}$. We assume that the channel matrix $\mathbf{H}_{p,t,n}$ is fixed during the single beam training period, i.e., $\mathbf{H}_{p,t} = \mathbf{H}_{p,t,i}$ $(i = 1, 2, , ..., N_c)$. Let $\mathbf{Y}_{p,t} = [\mathbf{y}_{p,t,1}, ..., \mathbf{y}_{p,t,N_c}]$, then

$$\mathbf{Y}_{p,t} = \mathbf{W}_{p,t}^H \mathbf{H}_{p,t} \mathbf{F}_t + \mathbf{W}_{p,t}^H \mathbf{N}_{p,t} \tag{16}$$

$$= \mathbf{W}_{p,t}^H \mathbf{H}_{p,t} \mathbf{F}_t + \mathbf{N}'_{p,t}, \tag{17}$$

where $\mathbf{F}_t = \begin{bmatrix} \mathbf{f}_{t,1} & \cdots & \mathbf{f}_{t,N_c} \end{bmatrix}$ and $\mathbf{N}'_{p,t} = \mathbf{W}_{p,t}^H \mathbf{N}_{p,t}$. Since the channel matrix $\mathbf{H}_{p,t}$ can be represented by a small number of parameters in the angular domain, we use the angular channel representation in (3), and as a result

$$\mathbf{Y}_{p,t} = \mathbf{W}_{p,t}^H \mathbf{A}_{p,t}^{(m)} \mathbf{\Lambda}_{p,t} (\mathbf{A}_{p,t}^{(b)})^H \mathbf{F}_t + \mathbf{N}'_{p,t} \tag{18}$$

$$= \sum_{l=1}^{L} \alpha_{p,t,l} \mathbf{W}_{p,t}^H \mathbf{a}^{(m)}(\theta_{p,t,l}^{(m)}) \left(\mathbf{a}^{(b)}(\theta_{p,t,l}^{(b)}) \right)^H \mathbf{F}_t + \mathbf{N}'_{p,t} \tag{19}$$

Note that the subscript n is dropped from the channel parameters $\theta_{p,t,l}^{(b)}$, $\theta_{p,t,l}^{(m)}$, and $\alpha_{p,t,l}$ since we assume that the CSI does not change during the beam training period. From (19), one can see that the channel estimation problem can be readily expressed as the joint estimation of the parameters $\alpha_{p,t,1}, ..., \alpha_{p,t,L}$, $\theta_{p,t,1}^{(b)}, ..., \theta_{p,t,L}^{(b)}$, and $\theta_{p,t,1}^{(m)}, ..., \theta_{p,t,L}^{(m)}$. In estimating these parameters, one can think of maximum likelihood estimation

$$\arg\max \; Pr(\mathbf{Y}_{p,t} | \alpha_{p,t,1}, ..., \alpha_{p,t,L}, \Theta_{p,t}^{(b)}, \Theta_{p,t}^{(m)}). \tag{20}$$

where $\Theta_{p,t}^{(b)} = \theta_{p,t,1}^{(b)}, ..., \theta_{p,t,L}^{(b)}$, and $\Theta_{p,t}^{(m)} = \theta_{p,t,1}^{(m)}, ..., \theta_{p,t,L}^{(m)}$. Due to the nonlinearity of AoD and AoA with respect to the received vector, joint parameter estimation tends to be computationally infeasible. Even though the optimization is performed over the discretized parameter space, it requires huge computational complexity for $L > 1$. Alternatively, one can use the virtual channel representation, $\mathbf{H}_{p,t} = \mathbf{A}^{(m)} \mathbf{H}_{p,t,n}^{(v)} \left(\mathbf{A}^{(b)} \right)^H$ in (17) and construct the observation vector $\mathbf{y}_{p,t}$ by

Algorithm 1 OMP algorithm to estimate mmWave channel

Initialize : $\mathbf{r}^{(0)} = \mathbf{y}_{p,t}, \Omega_0 = \{\}$

Output : Support set Ω and channel gain $\hat{\mathbf{h}}_\Omega$

1: **for** $i = 1$ to L... **do**

2: Select the index maximizing the magnitude of the inner product between ψ_i and $\mathbf{r}^{(i-1)}$:

$$i^* = \arg\max_i \ |\psi_i^H \mathbf{r}^{(i-1)}|^2,$$

where ψ_i is the ith column of $\mathbf{\Phi}_{p,t}$.

3: Update the support set Ω:

$$\Omega_i = \Omega_{i-1} \cup i^*$$

4: Update the signal amplitude $\hat{h}_\Omega$ and the residual $r^{(i)}$;

$$\hat{\mathbf{h}}_{\Omega_i} = \left((\mathbf{\Phi}_{p,t})_{\Omega_i}^H (\mathbf{\Phi}_{p,t})_{\Omega_i}\right)^{-1} (\mathbf{\Phi}_{p,t})_{\Omega_i}^H \mathbf{y}_{p,t}$$
$$\mathbf{r}^{(i)} = \mathbf{y}_{p,t} - (\mathbf{\Phi}_{p,t})_{\Omega_i} \hat{\mathbf{h}}_{\Omega_i}$$

5: **end for**

stacking the columns of $\mathbf{Y}_{p,t}$ as [9]

$$\begin{aligned}
\mathbf{y}_{p,t} =& \mathrm{vec}(\mathbf{Y}_{p,t}) \\
=& \mathrm{vec}(\mathbf{W}_{p,t}^H \mathbf{A}^{(m)} \mathbf{H}_{p,t}^{(v)} \left(\mathbf{A}^{(b)}\right)^H \mathbf{F}_t + \mathbf{N}'_{p,t}) \\
=& \left(\mathbf{F}_t^T \left(\mathbf{A}^{(b)}\right)^*\right) \otimes \left(\mathbf{W}_{p,t}^H \mathbf{A}^{(m)}\right) \mathrm{vec}\left(\mathbf{H}_{p,t}^{(v)}\right) + \mathrm{vec}(\mathbf{N}'_{p,t}) \\
=& \mathbf{\Phi}_{p,t} \mathbf{h}_{p,t} + \mathbf{n}_{p,t}, \qquad (21)
\end{aligned}$$

where $\mathbf{\Phi}_{p,t} = \left(\mathbf{F}_t^T \left(\mathbf{A}^{(b)}\right)^*\right) \otimes \left(\mathbf{W}_{p,t}^H \mathbf{A}^{(m)}\right)$, $\mathbf{h}_{p,t} = \mathrm{vec}\left(\mathbf{H}_{p,t}^{(v)}\right)$, and $\mathbf{n}_{p,t} = \mathrm{vec}\left(\mathbf{N}'_{p,t}\right)$. Note that the channel estimation is equivalent to the estimation of the unknown vector $\mathbf{h}_{p,t}$ from the received vector $\mathbf{y}_{p,t}$ in (21). In practical scenarios, we cannot use as many training beams as desired due to the limitation of resources, which in turn means that the number of rows in $\mathbf{\Phi}_{p,t}$ should be less than the number of columns. While obtaining an accurate estimate of $\mathbf{h}_{p,t}$ in an underdetermined systems is in general very difficult, we can accurately recover $\mathbf{h}_{p,t}$ by exploiting the sparsity of the channel vector $\mathbf{h}_{p,t}$ [26]. In estimating the channel $\mathbf{h}_{p,t}$, we can basically use

any sparse recovery algorithm including orthogonal matching pursuit (OMP) [27] described in Algorithm 1.

III. Proposed Dedicated Beam Training For Mobility Scenario

In this section, we present the new beam training technique to support the users under mobility. Since the common beam training should serve all users in a cell, a large number of beams to cover wide range of direction are needed. As mentioned, when there are moving users or the positions of devices change, a period for the common beam training should be reduced, thereby increasing training overhead. Instead of using the beam cycling, we use the *dedicated beam* targeted for the users under high mobility. Using the information on the target user's channel, we can enhance the efficiency of the beam training employing only a small number of beams optimized for the specific user. In the next subsections, we present the overall beam training protocol, dynamic channel characteristics, optimal beam design, and new channel estimation method.

A. Description of Overall Beam Training Protocol

Fig. 4 depicts the proposed beam training protocol. For the users under mobility (e.g., user 3, 4, and 5 in Fig. 4), the basestation transmits the dedicated beams along with the common beams. As shown in Fig. 4, the basestation sends N_d training beams for the dedicated beam training where the number of dedicated beams N_d is much smaller than the number of common beams N_c. In order to serve high mobility users, the dedicated beams are transmitted more frequently than the common beams, i.e., $T_d < T_c$. As shown in Fig. 4, the periodicity of the dedicated beam can be adjusted depending on the extent of mobility. The proposed dedicated beam training involves the interaction between the base-station and the target user. Based on the received beams, the target user estimates the CSI and pick the optimal beams to be used by the base-station in the next beam training period. The selected beam indices are fed back to the base-station and the same beam training procedure repeats until the frame ends. Since the beamforming vectors are optimally selected based on the CSI of the user, the good channel estimation quality and the beamforming accuracy can be maintained even with a small number of beam transmissions.

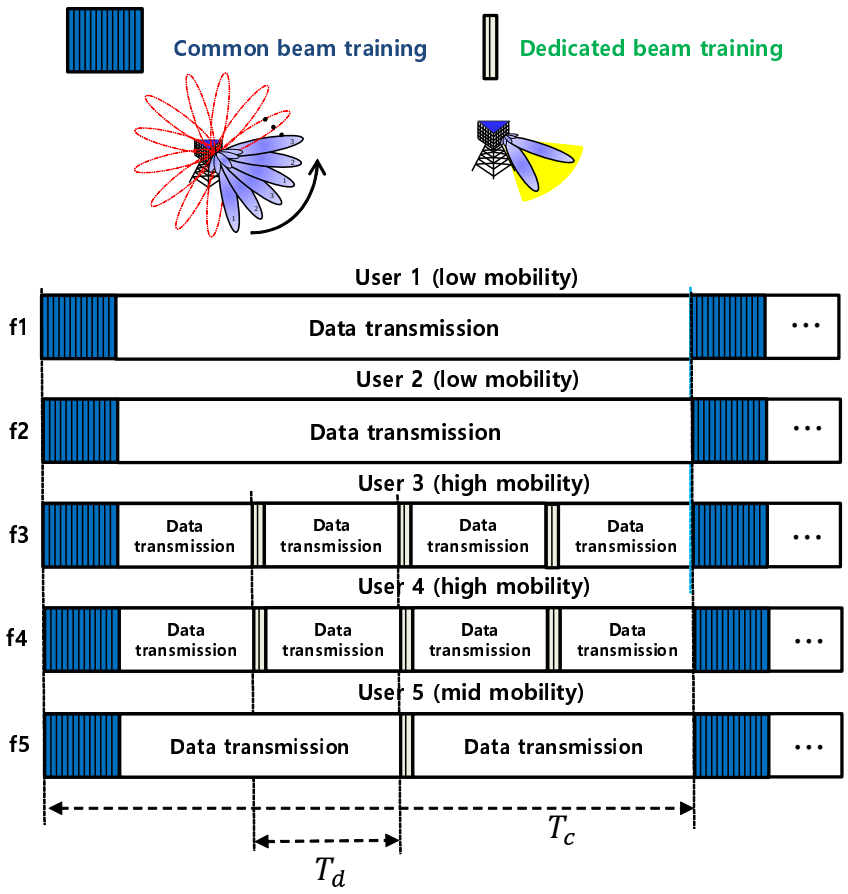


Fig. 4. Frame structure of the proposed dedicated beam training

B. Dynamic mmWave Channel Characteristics

In this subsection, we describe the statistical characteristics of mmWave channel under the mobility scenario. Under the condition that the channel is correlated in time and varying slowly, we can readily model this slowly-varying channel as a Markov process. Using the discrete representation of the AoA and AoD in (7), we model the AoD and AoA as the discrete state Markov process [30], [31]. The discrete state Markov process is described by the transition probability given by

$$T^{(b)}(m|n) = P\left(\theta^{(b)}_{p,t,l} = -1 + 2\frac{m-1}{M_b} \middle| \theta^{(b)}_{p,t-1,l} = -1 + 2\frac{n-1}{M_b}\right)$$
$$T^{(m)}(m|n) = P\left(\theta^{(m)}_{p,t,l} = -1 + 2\frac{m-1}{M_m} \middle| \theta^{(m)}_{p,t-1,l} = -1 + 2\frac{n-1}{M_m}\right). \quad (22)$$

In fact, one simple example of the transition probability is

$$T^{(b)}(m|n) = \frac{1}{C}\exp(-|m-n|^2/\sigma_l^2), \tag{23}$$

where C is the normalization constant and σ_l indicates the mobility of the user. This transition probability decays exponentially with the difference between the present and past AoD. Note that one can estimate the parameter σ_l from the user's mobility level (e.g. moving speed).

C. Optimal Beam Design

During the dedicated beam training period, the basestation needs to adapt the beamforming vectors based on user's CSI. In our work, since the dedicated beam training is performed in the angular space, we use the channel information represented in terms of the AoD and AoA. First, the basestation selects N_d beamforming vectors $(\mathbf{f}_{t,1}, ..., \mathbf{f}_{t,N_d})$ minimizing the channel estimation error of the target user from the beam codebook $\mathcal{D}$. We assume that the beam codebook $\mathcal{D} = \{\mathbf{d}_1, ..., \mathbf{d}_R\}$ contains the R pre-calculated beamforming vectors that correspond to the steering vectors for different directions and beam width[1]. Then, the optimal beam selection problem can be formulated as

$$\left(\mathbf{f}_{t,1}^*, ..., \mathbf{f}_{t,N_d}^*\right) = \underset{\mathbf{f}_{t,1},...,\mathbf{f}_{t,N_d}\in\mathcal{D}}{\arg\min}\, E\left[\|\mathbf{h}_{p,t} - \hat{\mathbf{h}}_{p,t}\|_2^2\right], \tag{24}$$

where $\hat{\mathbf{h}}_{p,t}$ is the estimate of $\mathbf{h}_{p,t}$. Since the channel estimation performance is determined by both the combining vector $\mathbf{w}_{p,t}$ and the beamforming vector $\mathbf{f}_t$, they should be jointly determined. In order to simplify the optimization process, we decouple joint optimization step by using the ideal combining matrix $\mathbf{W}_{p,t} = \mathbf{A}_{p,t}^{(m)}$ in the system model and optimizing the cost function only with respect to the beamforming vectors $\mathbf{F}_t = [\mathbf{f}_{t,1}, ..., \mathbf{f}_{t,N_d}]$. After the support $\mathcal{S}_h$ in $\mathbf{h}_{p,t}$, (corresponding to the active AoD bins) is estimated, the channel gain can be estimated by the linear projection of $\mathbf{y}_{p,t}$ onto the subspace spanned by $\mathcal{S}_h$. Since the channel gain is calculated by the function of AoD, the channel estimation performance is determined by the AoD estimation accuracy. Thus, the beam selection problem is reformulated as

$$\mathbf{f}_{t,1}^*, ..., \mathbf{f}_{t,N_d}^* = \underset{\mathbf{f}_{t,1},...,\mathbf{f}_{t,N_d}\in\mathcal{D}}{\arg\min}\, E\left[\sum_{l=1}^{L}\left|\theta_{p,t,l}^{(b)} - \hat{\theta}_{p,t,l}^{(b)}\right|^2\right], \tag{25}$$

[1] The beam-width can be increased by turning off the subset of the antennas

where $\hat{\theta}_{p,t,l}^{(b)}$ is the estimate of $\theta_{p,t,l}^{(b)}$. While the cost function in (25) is a reasonable choice, since mmWave channel estimation is highly related to the AoD estimation, it depends on channel estimation algorithm. To avoid the dependency of the cost function in (25) on the algorithm selection, we use the analytic bound of AoD variance as a performance metric. Specifically, we derive the lower bound of $E\left[\sum_{l=1}^{L}\left|\theta_{p,t,l}^{(b)}-\hat{\theta}_{p,t,l}^{(b)}\right|^2\right]$ using information inequality [28]. Let $\underline{\theta}=\left[\theta_{p,t,1}^{(b)},...,\theta_{p,t,L}^{(b)},\alpha_{p,t,1},...,\alpha_{p,t,L}\right]^T$, then the Cramer-Rao Lower bound (CRLB) for the parameter $\underline{\theta}$ is given by

$$\mathrm{Cov}\left(\underline{\theta}\right)\succeq\mathbf{V}^{-1} \tag{26}$$

$$(\mathbf{V})_{i,j}=E\left[\frac{\partial\log p(\mathbf{Y}_{p,t}|\underline{\theta})}{\partial(\underline{\theta})_i}\frac{\partial\log p(\mathbf{Y}_{p,t}|\underline{\theta})}{\partial(\underline{\theta})_j}\right], \tag{27}$$

where $\mathbf{V}$ is the Fisher's information matrix and $(\mathbf{V}^{-1})_{l,l}$ is the CRLB of $\theta_{p,t,l}^{(b)}$ for $1\leq l\leq L$. In our setup, the CRLB is a function of the deterministic parameters $\theta_{p,t,1}^{(b)},...,\theta_{p,t,L}^{(b)}$ so that we cannot determine the CRLB without the knowledge of them. As heuristic surrogate, we use the CRLB weighted with respect to the probability distribution $Pr(\theta_{p,t,1}^{(b)},...,\theta_{p,t,L}^{(b)})$ as a cost function, i.e.,

$$\begin{aligned}
C_{avg}(\mathbf{f}_{t,1},...,\mathbf{f}_{t,N_d}) =& E_{\theta_{p,t,1}^{(b)},...,\theta_{p,t,L}^{(b)}}\left[\sum_{l=1}^{L}[\mathrm{Cov}\left(\underline{\theta}\right)]_{l,l}\right]\\
\geq& \sum_{\theta_{p,t,1}^{(b)},...,\theta_{p,t,L}^{(b)}}\left[\sum_{l=1}^{L}(\mathbf{V}^{-1})_{l,l}\right]\\
=& \sum_{\theta_{p,t,1}^{(b)},...,\theta_{p,t,L}^{(b)}}\left(\sum_{l=1}^{L}(\mathbf{V}^{-1})_{l,l}\right)Pr(\theta_{p,t,1}^{(b)},...,\theta_{p,t,L}^{(b)})\\
=& \sum_{m_1,...,m_L}\left(\sum_{l=1}^{L}(\mathbf{V}^{-1})_{l,l}\right)\prod_{i=1}^{L}Pr\left(\theta_{p,t,i}^{(b)}=-1+2\frac{m_i-1}{M_b}\right).
\end{aligned} \tag{28}$$

where the AoDs for different paths are assumed to be statistically independent in (28). The distribution of the AoD is obtained by the proposed channel estimator, which will be described later. Specifically, based on the training beams received during the previous beam training periods, the proposed channel estimator predicts the distribution of the AoD for the next beam training period. By doing so, we can account for the uncertainty on the AoD estimate in our beam selection strategy. Since the cost function is expressed as a function of N_d beamforming vectors

$\mathbf{f}_{t,1}, ..., \mathbf{f}_{t,N_d}$, the beam selection rule can be readily expressed as

$$\left(\mathbf{f}_{t,1}^*, ..., \mathbf{f}_{t,N_d}^*\right) = \underset{\mathbf{f}_{t,1},...,\mathbf{f}_{t,N_d} \in \mathcal{D}}{\arg\min} C_{avg}(\mathbf{f}_{t,1}, ..., \mathbf{f}_{t,N_d}). \tag{29}$$

D. Beam Selection with Single Path Scenario

First, we consider the beam selection rule for the single path scenario (i.e., $L = 1$) where the energy of single path is predominant. In this scenario, the beam selection rule in (25) can be simplified to

$$\left(\mathbf{f}_{t,1}^*, ..., \mathbf{f}_{t,N_d}^*\right) = \underset{\mathbf{f}_{t,1},...,\mathbf{f}_{t,N_d} \in \mathcal{D}}{\arg\min} E\left|\theta_{p,t,1}^{(b)} - \hat{\theta}_{p,t,1}^{(b)}\right|^2. \tag{30}$$

Also, since $L = 1$, the received signal vector $\mathbf{Y}_{p,t}$ in (19) is simplified to

$$\mathbf{Y}_{p,t} = \alpha_{p,t,1}\mathbf{W}_t^H \mathbf{a}^{(m)}(\theta_{p,t,1}^{(m)})\left(\mathbf{a}^{(b)}(\theta_{p,t,1}^{(b)})\right)^H \mathbf{F}_t + \mathbf{N}_t' \tag{31}$$

Using the ideal combining matrix $\mathbf{W}_t = \mathbf{a}^{(m)}(\theta_{p,1,t}^{(m)})$, we have

$$\mathbf{Y}_{p,t} = \alpha_{p,t,1}\left(\mathbf{a}^{(b)}(\theta_{p,t,1}^{(b)})\right)^H \mathbf{F}_t + \mathbf{N}_t' \tag{32}$$

Since $\mathbf{Y}_{p,t}$ is a row vector, we can rewrite (32) as

$$\mathbf{y}_{p,t} = \mathbf{Y}_{p,t}^T = \mathbf{F}_t^T \mathbf{a}^{(b)}(\theta_{p,t,1}^{(b)})^* \alpha_{p,t,1} + \mathbf{n}_{p,t}, \tag{33}$$

and one can show that the CRLB of $\theta_{p,t,1}^{(b)}$ is given by (see Appendix A)

$$\begin{aligned} Var(\theta_{p,t,1}^{(b)}) &\geq (\mathbf{V}^{-1})_{3,3} \\ &= [(\mathbf{V})_{3,3} - 2Re\{(\mathbf{V})_{3,1}((\mathbf{V})_{1,1})^{-1}(\mathbf{V})_{1,3}\}]^{-1} \end{aligned} \tag{34}$$

where

$$\begin{aligned} (\mathbf{V})_{3,3} &= \frac{1}{\sigma^2}\left\|\alpha_{p,t,1}\mathbf{F}_t^T \frac{\partial(\mathbf{a}^{(b)}(\theta_{p,t,1}^{(b)})^*)}{\partial\theta_{p,t,1}^{(b)}}\right\|^2, \\ (\mathbf{V})_{1,3} &= (\mathbf{V})_{3,1}^* = \frac{1}{2\sigma^2}\alpha_{p,t,1}\frac{\partial\mathbf{a}^{(b)}(\theta_{p,t,1}^{(b)})}{\partial\theta_{p,t,1}^{(b)}}\mathbf{F}_t^*\mathbf{F}_t^T\mathbf{a}^{(b)}(\theta_{p,t,1}^{(b)})^* \\ (\mathbf{V})_{1,1} &= \frac{1}{2\sigma^2}\left\|\mathbf{F}_t^T\mathbf{a}^{(b)}(\theta_{p,t,1}^{(b)})^*\right\|^2. \end{aligned}$$

By taking similar step to (28), the average cost function can be obtained by weighting the lower

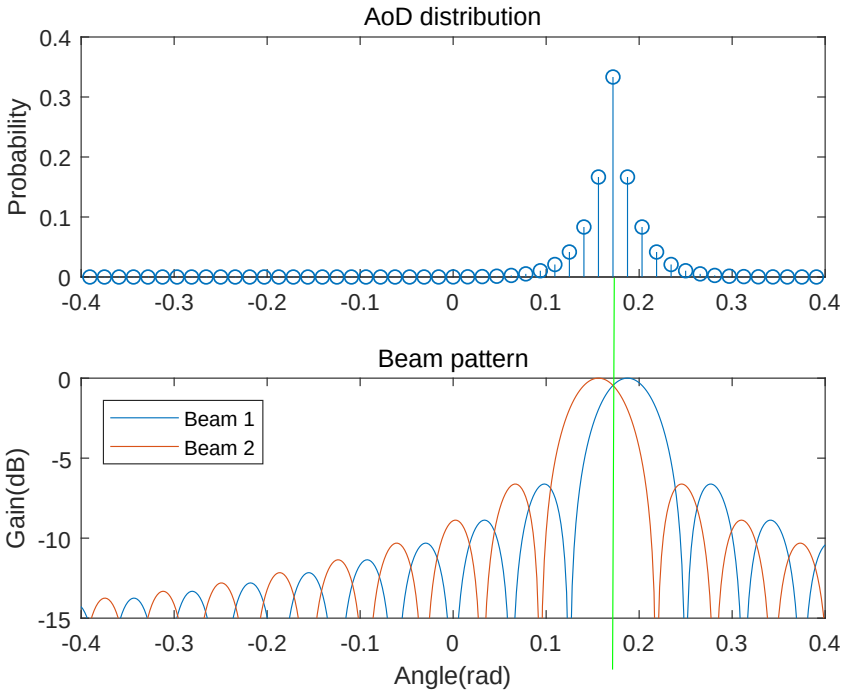


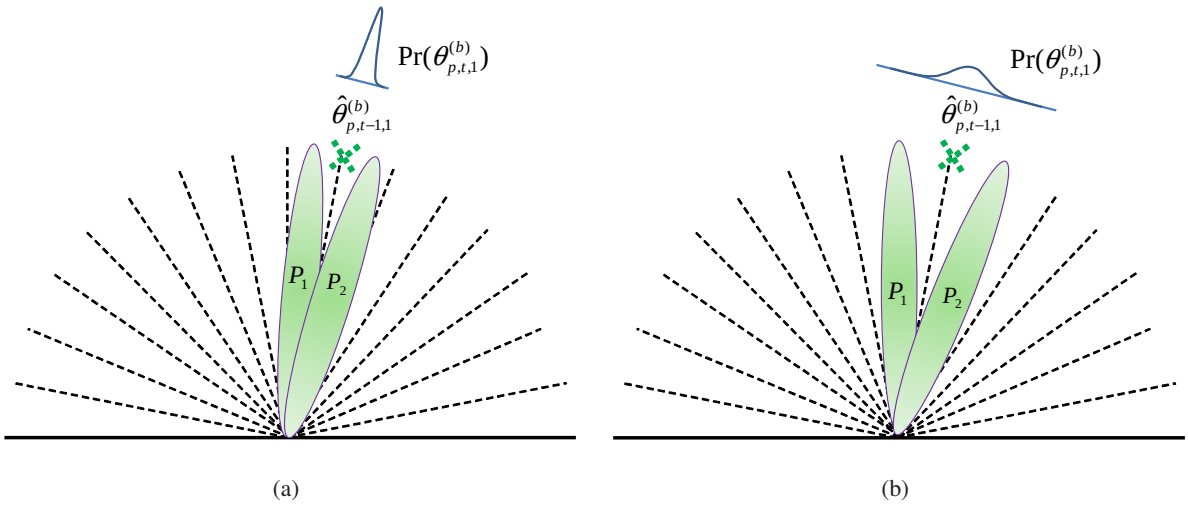
Fig. 5. The beam patterns found by the proposed beam selection for the given AoD distribution. Notice the symmetry of the two beam patterns with respect to the center of the distribution.

bound $(\mathbf{V}^{-1})_{3,3}$ with the distribution of $\theta_{p,t,1}^{(b)}$, i.e.,

$$C_{avg}(\mathbf{f}_{t,1},...,\mathbf{f}_{t,N_d}) = \sum_{m=1}^{M_b} (\mathbf{V}^{-1})_{3,3} Pr\left(\theta_{p,t,1}^{(b)} = -1 + 2\frac{m-1}{M_b}\right). \tag{35}$$

Then, we search for N_d beam indices from the beam codebook $\mathcal{D}$ that minimizes the cost function $C_{avg}(\mathbf{f}_{t,1},...,\mathbf{f}_{t,N_d})$. Although this process is combinatoric in nature and thus computationally burdensome, the computational complexity can be reduced significantly by considering only a small number of beams. In our work, we consider dual beam scenario ($N_d = 2$), which offers small search complexity as well as low training overhead. In the simulation section, we will demonstrate that the performance of the proposed method using two training beams is comparable to the common beam training using $N_c = 32$ beams. In addition, we will also show that using a single beam does not provide satisfactory performance. When $N_d = 2$, we search for two beam indices (i, j) from $\mathcal{D}$ as

$$(i^*, j^*) = \underset{(i,j)\in\mathcal{D}}{\arg\max}\ C_{avg}(\mathbf{f}_{t,1} = \mathbf{d}_i, \mathbf{f}_{t,2} = \mathbf{d}_j). \tag{36}$$



(a) (b)

Fig. 6. Example of the proposed beam selection for different AoD distributions.

To find out (i^*, j^*), we need to evaluate $\binom{M_b}{2}$ combinations of all beam indices. Fig. 5 shows the beam patterns obtained by the proposed selection criterion. When the AoD distribution is symmetric, directions of the optimal beam pair are also symmetric with respect to the center of the distribution. Exploiting this symmetry, we only have to determine the angle formed by two beamforming vectors. Then, two dimensional grid search can be reduced to the one dimensional search as

$$(i_0 + \delta^*, i_0 - \delta^*) = \underset{i_0+\delta, i_0-\delta \in \mathcal{D}}{\arg\max}\ C_{avg}(\mathbf{f}_{t,1} = \mathbf{d}_{i_0+\delta}, \mathbf{f}_{t,2} = \mathbf{d}_{i_0-\delta}), \tag{37}$$

where i_0 is the beam index corresponding to the center of the AoD distribution. Note that the beam index parameter δ^* corresponds to the optimal angle between two selected training beams. (see Fig. 6.)

E. Beam Selection With Multi-path Scenario

When there are multiple paths, we need to select beams based on the cost metric defined over the set of parameters $\underline{\theta} = [\alpha_{p,t,1}, \alpha^*_{p,t,1}, ..., \alpha_{p,t,L}, \alpha^*_{p,t,L}, \theta^{(b)}_{p,t,1}, .., \theta^{(b)}_{p,t,L}]$. As the number of parameters to be considered increases, the derivation of the CRLB would be cumbersome and also the optimization process requires huge computational complexity. To avoid this hassle, we use a simple beam selection method which is in essence an extension of the beam selection strategy for $L = 1$. The basic idea is to use the dual beams optimized for each path. In this

case, the basestation sends $2L$ beams in total. Since the number of paths L is small in mmWave environments, the training resources for the dedicated beam training is much smaller than the common beam training (i.e., $2L \ll N_c$). Since our channel estimator is designed to provide the distribution of AoD for each path, (i.e., $Pr(\theta^{(b)}_{p,t,1}), ..., Pr(\theta^{(b)}_{p,t,L})$), the optimal beam spacing δ^* is obtained by running the search algorithm for each path. Though such proposed per-path beam selection seems to be a bit heuristic, our numerical evaluation shows that it offers good system performance with reasonably small training overhead.

F. New Channel Estimation for Dedicated Beam Training

In order to estimate the channel $\mathbf{H}_{p,t}$ in mmWave communication systems, we first need to estimate the channel parameters $\alpha_{p,t,1}, ..., \alpha_{p,t,L}$, $\theta^{(b)}_{p,t,1}, ..., \theta^{(b)}_{p,t,L}$, and $\theta^{(m)}_{p,t,1}, ..., \theta^{(m)}_{p,t,L}$ from the received vector $\mathbf{y}_{p,t}$. Under the quantized channel model, ($\mathbf{y}_{p,t} = \mathbf{\Phi}_{p,t}\mathbf{h}_{p,t} + \mathbf{n}_{p,t}$), the AoD and AoA parameters correspond to the support of the channel vector $\mathbf{h}_{p,t}$ (i.e., the set of indices of nonzero elements in $\mathbf{h}_{p,t}$) and its amplitude. In this subsection, we propose a greedy sparse channel estimation algorithm that jointly estimates the support and the amplitude of $\mathbf{h}_{p,t}$ from $\mathbf{y}_{p,t}$. Note that the proposed channel estimator can also generate the distribution of the current AoD parameter $Pr(\theta^{(b)}_{p,t,l})$ using the measurement vectors $\mathbf{y}_{p,1}, ..., \mathbf{y}_{p,t-1}$ acquired until the $(t-1)$th beam training period.

Let $\mathcal{S}_{p,t} = \{s_{p,t,1}, ..., s_{p,t,L}\}$ be the support of the channel vector $\mathbf{h}_{p,t}$, where the support index $s_{p,t,l}$ corresponds to the pair of the AoD $\theta^{(b)}_{p,t,l}$ and the AoA $\theta^{(m)}_{p,t,l}$. Then, the received vector $\mathbf{y}_{p,t}$ can be expressed as

$$\mathbf{y}_{p,t} = (\mathbf{\Phi}_{p,t})_{\mathcal{S}_{p,t}}\mathbf{g}_{p,t} + \mathbf{n}_{p,t}. \tag{38}$$

where $\mathbf{g}_{p,t}$ is the $L \times 1$ vector containing nonzero gains in $\mathbf{h}_{p,t}$. Recall that $(\mathbf{X})_\Omega$ is the submatrix of $\mathbf{X}$ that contains columns indexed by Ω [2]. We assume that $\mathbf{g}_{p,t}$ is the Gaussian random vector $CN(0, \lambda\mathbf{I})$. Since the channel vector $\mathbf{h}_{p,t}$ is completely determined by two parameters, viz., the support $\mathcal{S}_{p,t}$ and the magnitude of the gain $\mathbf{g}_{p,t}$, the channel estimate can be obtained by finding the joint estimate of $\mathcal{S}_{p,t}$ and the amplitude $\mathbf{g}_{p,t}$. The joint maximum a posteriori (MAP) estimate

[2]For example, $\left(\begin{bmatrix} 1 & 2 & 3 & 4 \\ 4 & 3 & 2 & 1 \end{bmatrix}\right)_{\{2,4\}} = \begin{bmatrix} 2 & 4 \\ 3 & 1 \end{bmatrix}$.

of $\mathcal{S}_{p,t}$ and $\mathbf{g}_{p,t}$ is given by

$$\left(\hat{\mathcal{S}}_{p,t}, \hat{\mathbf{g}}_{p,t}\right) = \underset{\mathcal{S}_{p,t}, \mathbf{g}_{p,t}}{\arg\max} \ln Pr\left(\mathcal{S}_{p,t}, \mathbf{g}_{p,t} | \mathbf{y}_{p,1}, ..., \mathbf{y}_{p,t}\right). \tag{39}$$

Using $Pr(\mathcal{S}_{p,t}, \mathbf{g}_{p,t}|\mathcal{Y}_{p,t}) = Pr(\mathcal{S}_{p,t}|\mathcal{Y}_{p,t}) Pr(\mathbf{g}_{p,t}|\mathcal{S}_{p,t}, \mathcal{Y}_{p,t})$ and denoting $\mathcal{Y}_{p,t} = \{\mathbf{y}_{p,1}, ..., \mathbf{y}_{p,t}\}$, the MAP estimate $\hat{\mathcal{S}}_{p,t}$ of the support can be obtained from

$$\hat{\mathcal{S}}_{p,t} = \underset{\mathcal{S}_{p,t}}{\arg\max} \left[\max_{\mathbf{g}_{p,t}} \left(\ln Pr(\mathcal{S}_{p,t}|\mathcal{Y}_{p,t}) + \ln Pr(\mathbf{g}_{p,t}|\mathcal{S}_{p,t}, \mathcal{Y}_{p,t}) \right) \right] \tag{40}$$

$$= \underset{\mathcal{S}_{p,t}}{\arg\max} \left[\ln Pr(\mathcal{S}_{p,t}|\mathcal{Y}_{p,t}) + \max_{\mathbf{g}_{p,t}} \left(\ln Pr(\mathbf{g}_{p,t}|\mathcal{S}_{p,t}, \mathcal{Y}_{p,t}) \right) \right]. \tag{41}$$

Note that the channel amplitude vector $\mathbf{g}_{p,t}$ is Gaussian distributed when the support $\mathcal{S}_{p,t}$ is given, i.e., $Pr(\mathbf{g}_{p,t}|\mathcal{S}_{p,t}, \mathcal{Y}_{p,t}) \sim \mathcal{CN}(\bar{\mathbf{g}}_{p,t}, P_{p,t})$ where

$$\bar{\mathbf{g}}_{p,t} = \left((\mathbf{\Phi}_{p,t})^H_{\mathcal{S}_{p,t}} (\mathbf{\Phi}_{p,t})_{\mathcal{S}_{p,t}} + \frac{\sigma_n^2}{\lambda} I \right)^{-1} (\mathbf{\Phi}_{p,t})^H_{\mathcal{S}_{p,t}} \mathbf{y}_{p,t}$$

$$P_{p,t} = \sigma_n^2 \left((\mathbf{\Phi}_{p,t})^H_{\mathcal{S}_{p,t}} (\mathbf{\Phi}_{p,t})_{\mathcal{S}_{p,t}} + \frac{\sigma_n^2}{\lambda} I \right)^{-1}.$$

Since $\bar{\mathbf{g}}_{p,t}$ is the maximizer of $Pr(\mathbf{g}_{p,t}|\mathcal{S}_{p,t}, \mathcal{Y}_{p,t})$ in (41), we have

$$\hat{\mathcal{S}}_{p,t} = \underset{\mathcal{S}_{p,t}}{\arg\max} \left[\ln Pr(\mathcal{S}_{p,t}|\mathcal{Y}_{p,t}) - \ln \left(|\pi P_{p,t}| \right) \right]. \tag{42}$$

One can also show that the term $\ln Pr(\mathcal{S}_{p,t}|\mathcal{Y}_{p,t})$ in (42) can be expressed with respect to $Pr(\mathcal{S}_{p,t-1}|\mathcal{Y}_{p,t-1})$ in a recursive form as

$$\begin{aligned}
\ln Pr(\mathcal{S}_{p,t}|\mathcal{Y}_{p,t}) &= \ln Pr(\mathcal{S}_{p,t}|\mathcal{Y}_{p,t-1}, \mathbf{y}_{p,t}) \\
&= \ln Pr(\mathbf{y}_{p,t}|\mathcal{S}_{p,t}, \mathcal{Y}_{p,t-1}) + \ln Pr(\mathcal{S}_{p,t}|\,\mathcal{Y}_{p,t-1}) + C \\
&= -\ln \left| \pi \left(\sigma_n^2 I + \lambda (\mathbf{\Phi}_{p,t})^H_{\mathcal{S}_{p,t}} (\mathbf{\Phi}_{p,t})_{\mathcal{S}_{p,t}} \right) \right| \\
&\quad + \mathbf{y}^H_{p,t} (\mathbf{\Phi}_{p,t})_{\mathcal{S}_{p,t}} P_{p,t} (\mathbf{\Phi}_{p,t})^H_{\mathcal{S}_{p,t}} \mathbf{y}_{p,t} \\
&\quad + \ln \sum_{\mathcal{S}_{p,t-1}} Pr(\mathcal{S}_{p,t}|\mathcal{S}_{p,t-1}) Pr(\mathcal{S}_{p,t-1}|\mathcal{Y}_{p,t-1}) + C',
\end{aligned} \tag{43}$$

where C and C' are the terms unrelated to $\mathcal{S}_{p,t}$. Note that the conditional distribution $Pr(\mathcal{S}_{p,t}|\mathcal{S}_{p,t-1})$ is determined by the discrete-state Markov process described in (22). Plugging (43) into (42),

Algorithm 2 Proposed greedy recovery algorithm

Initialize : $\mathbf{r}^{(0)} = \mathbf{y}_{p,t}, \Omega_0 = \{\}$

Output : support set Ω, signal amplitude $\hat{x}_\Omega$, AoD distribution, $Pr(\mathcal{S}_{p,t+1} = \{\omega\}|\mathcal{Y}_{p,t})$

1: **for** $l = 1$ to L... **do**

2: Select the index ω that leads to the largest objective function:

$$\omega^{(i)} = \arg\max_{\omega} \ \psi(\omega)$$

3: Update the support set Ω_i

$$\Omega_i = \Omega_{i-1} \cup \omega^{(i)}$$

4: Update the signal amplitude set $\mathbf{x}_{\Omega_i}$ and the residual signal $\mathbf{r}_t^{(l)}$

$$\mathbf{x}_{\Omega_i} = \left((\mathbf{\Phi}_{p,t})_{\Omega_i}^H (\mathbf{\Phi}_{p,t})_{\Omega_i} + \frac{\sigma_n^2}{\lambda} I \right)^{-1} (\mathbf{\Phi}_{p,t})_{\Omega_i}^H \mathbf{y}_{p,t}$$

$$\mathbf{r}_t^{(l)} = \mathbf{y}_{p,t} - (\mathbf{\Phi}_{p,t})_{\Omega_i} \mathbf{x}_{\Omega_i}$$

5: Update the *a posteriori* distribution using (47).

6: **end for**

7: Calculate the AoD distribution for the next beam training period using (48)

we obtain the expression for the objective function as

$$\begin{aligned} \psi(\mathcal{S}_{p,t}) = &- \ln \left| \pi \left(\sigma_n^2 I + \lambda (\mathbf{\Phi}_{p,t})_{\mathcal{S}_{p,t}}^H (\mathbf{\Phi}_{p,t})_{\mathcal{S}_{p,t}} \right) \right| \\ &+ \mathbf{y}_{p,t}^H (\mathbf{\Phi}_{p,t})_{\mathcal{S}_{p,t}} P_{p,t} (\mathbf{\Phi}_{p,t})_{\mathcal{S}_{p,t}}^H \mathbf{y}_{p,t} \\ &+ \ln \sum_{\mathcal{S}_{p,t-1}} Pr(\mathcal{S}_{p,t}|\mathcal{S}_{p,t-1}) Pr(\mathcal{S}_{p,t-1}|\mathcal{Y}_{p,t-1}) - \ln \left(|\pi P_{p,t}| \right). \end{aligned} \tag{44}$$

To find out the MAP estimate of the support, we need to evaluate (44) for all possible combinations of $\mathcal{S}_{p,t}$. Since the search complexity growth exponentially with L, this option is infeasible in practice.

In this work, we propose a computationally efficient greedy algorithm that finds the candidate element of $\mathcal{S}_{p,t}$ in an iterative fashion. Adopting the greedy strategy, we propose a cost function to find a predominant support index. Let Ω_i be the set of the support indices found by the ith iteration, then we compute the residual signal $\mathbf{r}^{(i)}$ for the ith iteration by subtracting the

contribution of the already found support elements in Ω_i from the observation $\mathbf{r}^{(i)}$. That is,

$$\mathbf{r}^{(i)} = \mathbf{y}_{p,t} - (\mathbf{\Phi}_{p,t})_{\Omega_i}\mathbf{x}_{\Omega_i} \tag{45}$$

where

$$\mathbf{x}_{\Omega_i} = \left((\mathbf{\Phi}_{p,t})_{\Omega_i}^H(\mathbf{\Phi}_{p,t})_{\Omega_i} + \frac{\sigma_n^2}{\lambda}I\right)^{-1}(\mathbf{\Phi}_{p,t})_{\Omega_i}^H\mathbf{y}_{p,t}.$$

Note that we assume that the effect of the support indices in Ω_i is perfectly subtracted from the residual signal $\mathbf{r}^{(i)}$. Since there exist a dominant support element in the residual $\mathbf{r}^{(i)}$, we aim to find it from the residual $\mathbf{r}^{(i)}$. Assuming the single support candidate $\omega \notin \Omega_{i-1}$, the objective function in (44) becomes

$$\begin{aligned}\psi(\omega) = &- \ln\left|\pi\left(\sigma_n^2 I + \lambda(\mathbf{\Phi}_{p,t})_\omega^H(\mathbf{\Phi}_{p,t})_\omega\right)\right| + \frac{1}{\sigma_n^2}\frac{\left|(\mathbf{\Phi}_{p,t})_\omega^H\mathbf{r}^{(i-1)}\right|^2}{\|(\mathbf{\Phi}_{p,t})_\omega\|^2 + \frac{\sigma_n^2}{\lambda}} \\ &+ \ln\sum_v Pr\left(s_{p,t}^{(i)} = \omega | s_{p,t-1}^{(i)} = v\right) Pr\left(s_{p,t-1}^{(i)} = v|\mathcal{Y}_{p,t-1}\right) - \pi\frac{\sigma_n^2}{\|(\mathbf{\Phi}_{p,t})_\omega\|^2 + \frac{\sigma_n^2}{\lambda}I},\end{aligned} \tag{46}$$

where $s_{p,t}^{(i)}$ is the index of the support not in Ω_{i-1}. Note that while the first two terms of (46) are found in the cost function of OMP, the last two terms account for temporal correlations of the support elements. Note that $Pr\left(s_{p,t}^{(i)} = \omega | s_{p,t-1}^{(i)} = v\right)$ can be given by the transition probability in (23), i.e., $\frac{1}{C}\exp(-|v-\omega|^2/\sigma_l^2)$. In a similar way as in (43), the term $Pr\left(s_{p,t-1}^{(i)} = v|\mathcal{Y}_{p,t-1}\right)$ in (46) can be recursively calculated as

$$\begin{aligned}\ln Pr(s_{p,t}^{(i)} = \omega|\mathcal{Y}_{p,t}) = &\frac{1}{\sigma_n^2}\frac{\left|(\mathbf{\Phi}_{p,t})_\omega^H\mathbf{r}^{(i-1)}\right|^2}{\|(\mathbf{\Phi}_{p,t})_\omega\|^2 + \frac{\sigma_n^2}{\lambda}} \\ &+ \ln\sum_v Pr\left(s_{p,t}^{(i)} = \omega | s_{p,t-1}^{(i)} = v\right) Pr\left(s_{p,t-1}^{(i)} = v|\mathcal{Y}_{p,t-1}\right) \\ &- \ln\left|\pi\left(\sigma_n^2 I + \lambda(\mathbf{\Phi}_{p,t})_\omega^H(\mathbf{\Phi}_{p,t})_\omega\right)\right|.\end{aligned} \tag{47}$$

Further, using the result in (47), the distribution of the AoD for the next beam training period can be obtained as

$$Pr(s_{p,t+1}^{(i)} = \omega|\mathcal{Y}_{p,t}) = \sum_v Pr(s_{p,t+1}^{(i)} = \omega | s_{p,t}^{(i)} = v) Pr(s_{p,t}^{(i)} = v|\mathcal{Y}_{p,t}). \tag{48}$$

This distribution is used to perform the beam selection for the next beam training period. We summarize the proposed greedy algorithm in Algorithm 2.

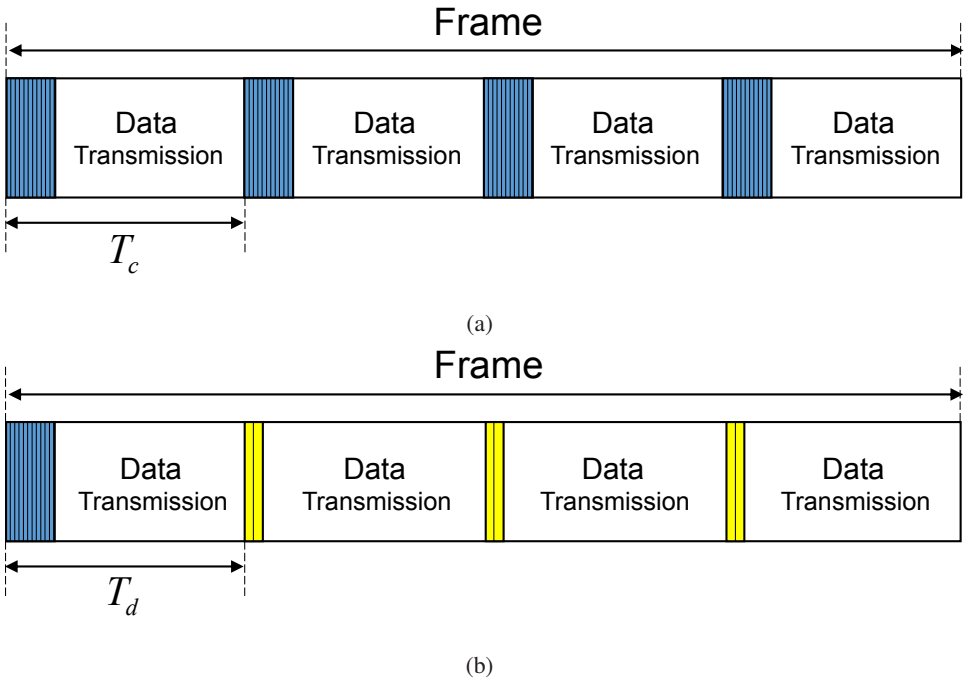


Fig. 7. Frame structure of (a) Conventional beam training and (b) dedicated beam training.

IV. Simulation Results and Discussions

A. Simulation Setup

In this section, we evaluate the performance of the proposed dedicated beam training technique. In our simulations, we assume 28GHz mmWave frequency band. We consider uniform linear arrays (ULAs) antenna whose adjacent antenna elements are spaced by the half wavelength. The numbers of the antenna elements in the basestation and the user are set to $N_b = N_m = 32$. Fig. 7 (a) and (b) present the frame structures for the conventional beam training and the proposed training used in our simulations. We assume that the symbol duration is 170μs. One frame consists of 10000 symbols in total, which include both training and data symbols. In the beginning of the frame, the beam cycling ($N_c = 32$ beam transmissions) is performed. In the conventional beam training, the beam cycling is repeated every T_c symbols. In the proposed beam training protocol, on the other hand, the dedicated beam training is performed every T_d symbols after the initial beam cycling. For the dedicated beam training, we use the dual beams (see Section III.D for details). The rest symbol slots (i.e., not used for beam training) are allocated for data transmission. The data is modulated by the binary phase shift keying (BPSK) modulation.

The data symbols are multiplied by the precoding matrix V obtained by the singular value decomposition of the channel estimate $\hat{\mathbf{H}}_{p,t} = U\Sigma V^H$. In the simulations, the channel matrix is generated according to the model described in Section II.B. We assume that the AoD and AoA change every symbol period according to the distribution $T^{(b)}(m|n) = \frac{1}{C}\exp(-|m-n|^2/\beta^2)$. Note that σ_l^2 in (23) can be converted from β^2 as

$$\sigma_l^2 = T_c\beta^2 \tag{49}$$

We also assume that the channel gain changes according to auto-regressive (AR) process given by

$$\alpha_n = \rho\alpha_{n-1} + v_n\sqrt{1-\rho^2} \tag{50}$$

where $v_n \sim N(0,1)$ and $\rho = 0.9999$. The resolution of the quantized AoD and AoA representation is $M_b = M_m = 128$. The beam codebook contains 128 beamforming vectors whose angles are uniformly spaced between $[-\pi/2, \pi/2]$. We generate 1000 frames (i.e., 10^7 symbols) to evaluate the bit error rate (BER). We define the training overhead as the ratio of the number of symbol slots used for training to the total number of symbols in a frame.

B. Simulation Results

First, we investigate how fast the AoD (or AoA) changes in time with respect to the parameter β. Fig. 8 shows the sample path of AoD (or AoA) in our channel model for different values of β. As β increases, the extent of mobility of users increases and thus AoD (or AoA) exhibits higher variations. In order to find how parameter β corresponds to the actual speed of a moving user, we generate the moving trajectory samples based on the second-order differential mobility model in [29] for the given speed of the user. Then, we obtain the maximum likelihood (ML) estimate of β from the trajectory samples. We see that $\beta = 0.5 \times 10^{-3}$ corresponds to the user moving at the speed of $3m/s$ $(= 10.8km/h)$ at 1 meter away from the base-station. Note that β scales linearly with the speed of the user.

Fig. 9 shows performance of the conventional beam cycling under mobility scenarios. We plot BER performance as a function of the training overhead for different mobility levels. We consider the multi-path channel with $L = 3$ and the signal to noise ratio (SNR) is set to 15 dB. When the training period T_c is set to 1000, the common beam training is performed only once

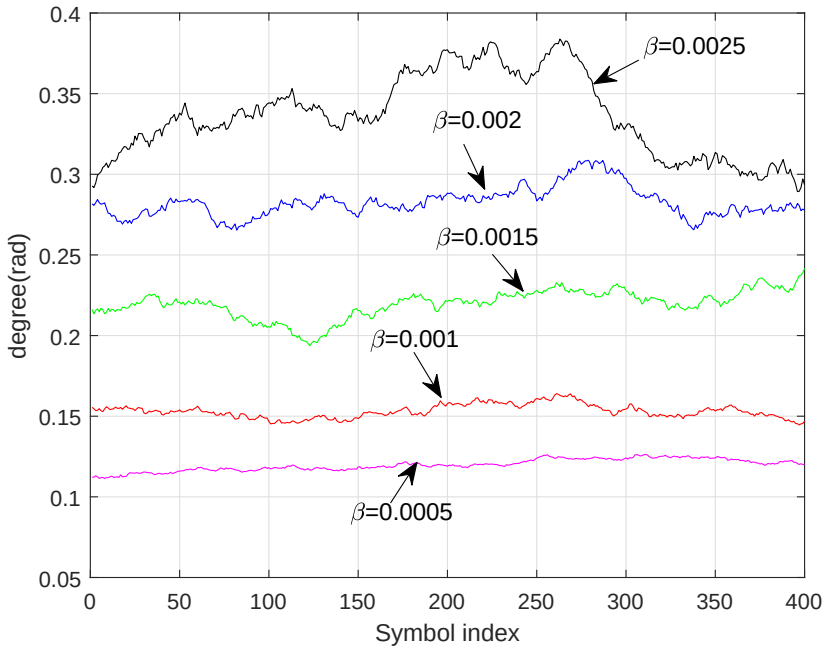


Fig. 8. Temporal variation of AoD or AoA for several values of β.

within the frame, resulting in the training overhead of 3.2%. As the training period T_c reduces to $500, 200$ and 50, the training overhead increases up to 6.4%, 16%, and 64%, respectively. As expected, the BER performance degrades as the extent of mobility for the users increases. To achieve better performance, the common beam training should be performed more frequently, leading to higher training overhead. For example, at the mobility of $\beta = 0.002$, the system should use more than 50% training overhead to achieve 10^{-2} BER. This results clearly demonstrate that the conventional beam training is inefficient for the mobility scenarios.

We next evaluate the performance of several beam training strategies. The following beam training schemes are considered in our experiments

1) Conventional beam cycling + OMP (BC + OMP): Beam cycling with $N_c = 32$ is performed with periodicity T_c. The AoD and AoA are estimated by OMP algorithm.
2) Conventional Kalman-based beam tracking [21] (Kalman BT): After initial beam cycling, the single beam is transmitted with period T_d. The AoD and AoA are estimated by the extended Kalman filter.
3) Dedicated dual beam training + OMP (DBT-dual + OMP): Proposed dual beam transmis-

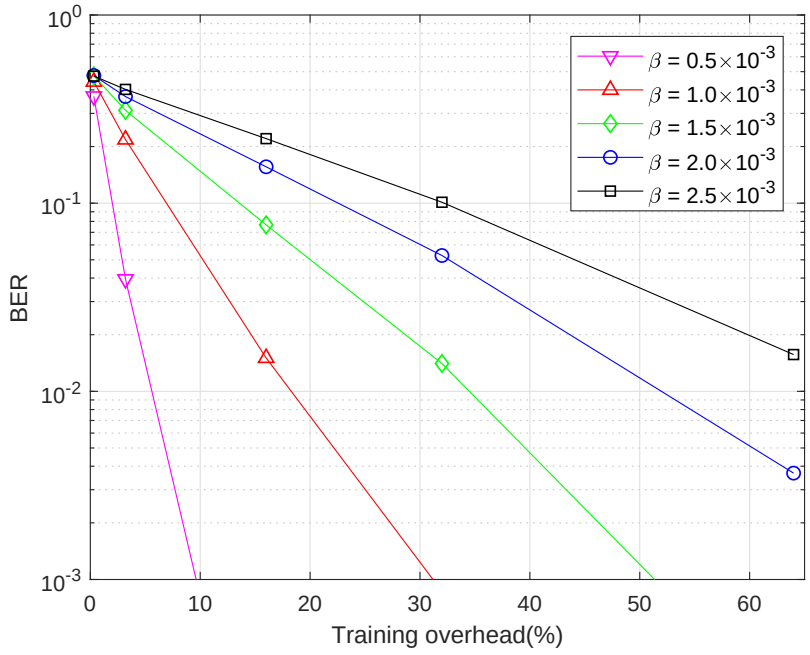
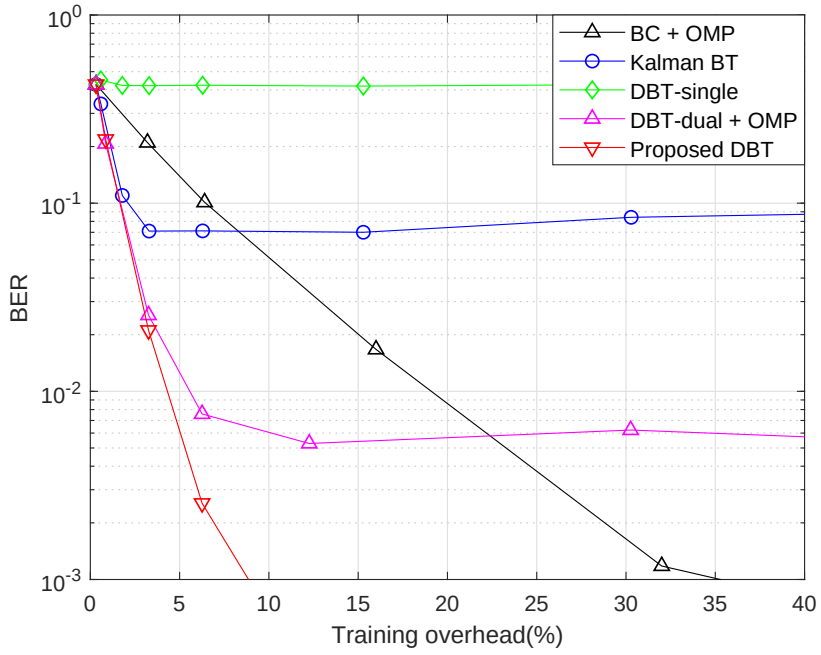


Fig. 9. Plot of BER versus training overhead for the conventional beam cycling.

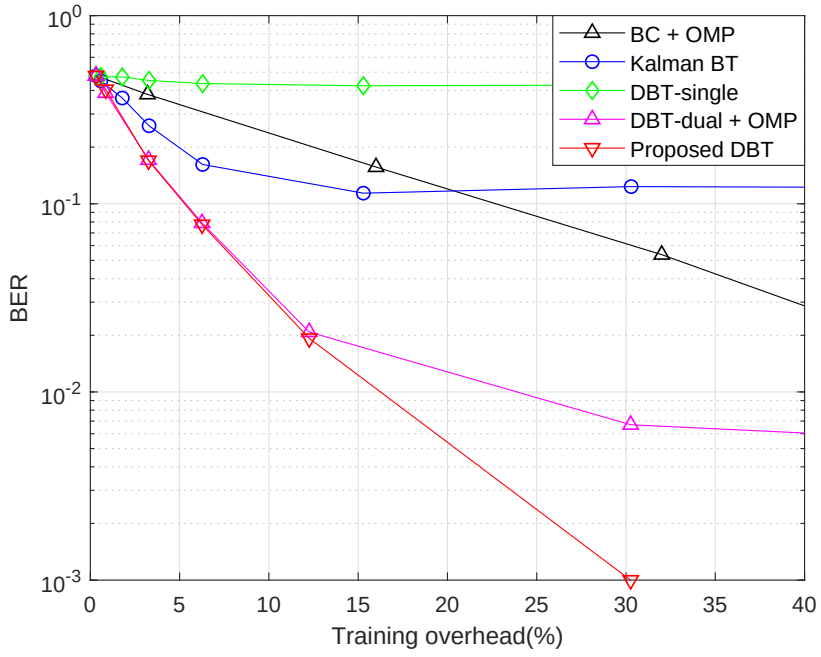
sion is used in combination with OMP algorithm. This algorithm is included to verify the superiority of the proposed channel estimator.

4) Dedicated single beam training (DBT-single): The single beam transmission (with periodicity T_d) is used in combination with the proposed greedy recovery algorithm. This algorithm is included to verify the effectiveness of the dual beam transmission.
5) Proposed dedicated dual beam training (Proposed DBT)

Fig. 10 provides the BER performance as a function of the training overhead for the several beam training schemes. By changing the periods of the beam training, T_c and T_d, we change the training overhead. We set $\beta = 0.001$ and 0.002 in Fig. 10 (a) and (b), respectively. The rest of the system parameters are the same as those used in Fig. 9. We observe that the proposed scheme outperforms the competing schemes with the same training overhead. Though both the proposed and conventional beam cycling schemes can achieve the BER level lower than 10^{-3}, the proposed scheme requires significantly less training overhead due to the short beam transmission period. Specifically, when $T_d = 1000, 200$ and 50, the training overheads of the proposed method are 0.9%, 3.3%, and 12.3%, respectively. This is in contrast to the conventional beam cycling



(a)

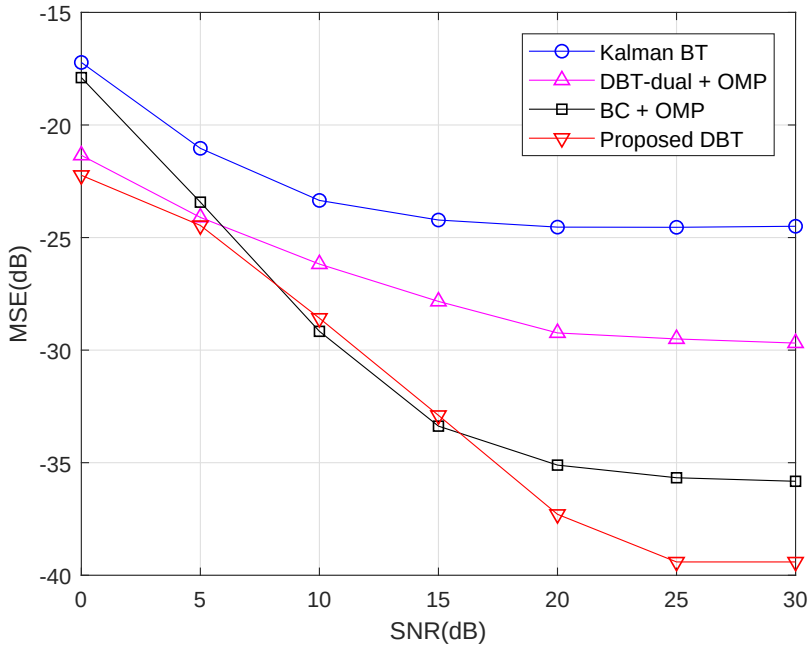


(b)

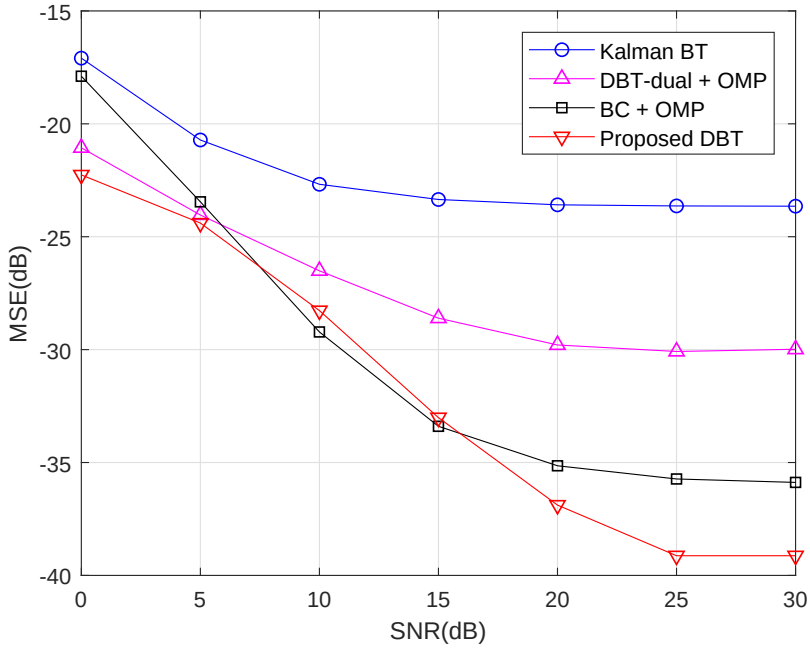
Fig. 10. Plot of BER versus training overhead for the dedicated beam training for (a) $\beta = 0.001$ and (b) $\beta = 0.002$.

requiring 6.4%, 16%, and 64% overhead for the same set of training period. Note that the training overhead of the proposed scheme can be reduced further when the beam transmission for the proposed scheme is performed only over the bandwidth occupied by the dedicated user. Next, we compare the proposed scheme with the single beam transmission approaches. Though the single beam transmission scheme has lower training overhead, they suffer from the performance loss. As shown in Fig. 10, with single beam transmission, the proposed channel estimator (not to mention OMP) does not work at all in this scenario. We observe that the system matrix becomes rank-deficient with single beam transmission and thus the support cannot be accurately estimated. The Kalman-based beam tracking scheme [21] performs slightly better than OMP but it performs much worse than the proposed method. The proposed scheme also outperforms the "DBT-dual + OMP" scheme due to the improved channel estimation accuracy.

We next evaluate the performance of the proposed channel estimation algorithm. We compare the proposed algorithm with two versions of OMP-based channel estimator (used for the beam cycling and the dedicated beam training) and the extended Kalman filter. In order to restrict our attention to the channel estimation performance only, the beam training periods T_c and T_d are equally set. For example, with $T_c = T_d = 100$, the training overhead for the beam cycling, dedicated dual beam training, and dedicated single beam training becomes 32%, 6%, and 3%, respectively. We fix the mobility parameter β to 0.001. The normalized mean square error (MSE) for channel estimation is measured only in the beam training period. Fig. 11 shows the MSE performance as a function of SNR. Note that the floor in the MSE for high SNR range is due to variation of the channels over the beam training period. When the dedicated beam training is performed, the proposed channel estimator achieves better MSE performance than OMP over the whole SNR range of interest. Note that this gain comes from the fact that the proposed algorithm exploits the temporal channel correlation across multiple beam training periods while OMP independently processes for each beam training period. We also observe that the proposed algorithm outperforms the extended Kalman filter-based channel estimator [21] as well. The proposed scheme achieves the performance gain over the beam cycling except for slight performance loss in mid SNR range. Combining the measurements across multiple beam training periods effectively, the proposed channel estimator achieves low MSE in spite of the 16-fold reduction in training overhead as compared to the conventional beam cycling.



(a)



(b)

Fig. 11. Channel estimation MSE versus SNR: (a) $T_c = T_d = 100$ and (b) $T_c = T_d = 200$.

V. Conclusions

In this paper, we have presented the novel beam training protocol that can significantly reduce training overhead for the mobility scenario in mmWave communications. The proposed dedicated beam training is designed to perform beam training only for high mobility users. Using a small number of training beams, the proposed dedicated beam training yields small training overhead. The proposed method maintains good system performance by employing the optimal beam selection strategy that finds the best beams in terms of channel estimation performance for the target user. In addition, we have devised the greedy sparse channel estimation algorithm that can exploits temporal channel correlations for the target user. Our numerical evaluations demonstrated that the proposed scheme achieves satisfactory system performance with small amount of training overhead as compared to the conventional methods.

VI. Appendix

A. Derivation of CRLB

In this subsection, we derive the CRLB for $\theta_{p,t,1}^{(b)}$. Since the channel parameters $\alpha_{p,t,1}$ and $\theta_{p,t,1}^{(b)}$ are unknown, we use the information inequality for the complex variable $\alpha_{p,t,1}$ and the real variable $\theta_{p,t,1}^{(b)}$. If we let $\underline{\theta} = [\alpha_{p,t,1}, \alpha_{p,t,1}^*, \theta_{p,t,1}^{(b)}]$, the CRLB of the real variable $\theta_{p,t,1}^{(b)}$ is given by [25]

$$CRLB = (\mathbf{V}^{-1})_{3,3} \tag{51}$$

$$= [(\mathbf{V})_{3,3} - 2Re\{(\mathbf{V})_{3,1}((\mathbf{V})_{1,1})^{-1}(\mathbf{V})_{1,3}\}]^{-1} \tag{52}$$

where $\mathbf{V}$ is Fisher information matrix. Recall that the (i,j)th element of $\mathbf{V}$, $(\mathbf{V})_{i,j}$ is given by

$$(\mathbf{V})_{i,j} = E\left[\frac{\partial \log p(\mathbf{y}_{p,t}|\underline{\theta})}{\partial(\underline{\theta})_i}\frac{\partial \log p(\mathbf{y}_{p,t}|\underline{\theta})}{\partial(\underline{\theta})_j}\right] \tag{53}$$

$$= -E\left[\frac{\partial}{\partial(\underline{\theta})_i^*}\frac{\partial \log p(\mathbf{y}_{p,t}|\underline{\theta})}{\partial(\underline{\theta})_j}\right]. \tag{54}$$

In addition, the ith diagonal element of $\mathbf{V}$, $(\mathbf{V})_{i,i}$ is given by

$$(\mathbf{V})_{i,i} = E\left[\left(\frac{\partial \log p(\mathbf{y}_{p,t}|\underline{\theta})}{\partial(\underline{\theta})_i}\right)^2\right] = -E\left[\frac{\partial^2 \log p(\mathbf{y}_{p,t}|\underline{\theta})}{\partial(\underline{\theta})_i^2}\right]. \tag{55}$$

Now, we derive the expression for $(\mathbf{V})_{1,1}$, $(\mathbf{V})_{3,3}$, $(\mathbf{V})_{3,1}$, and $(\mathbf{V})_{1,3}$. First, $(\mathbf{V})_{3,3}$ is obtained as

$$(\mathbf{V})_{3,3} = -E\left[\frac{\partial^2 \log p(\mathbf{y}_{p,t}|\underline{\theta})}{\partial\left(\theta_{p,t,1}^{(b)}\right)^2}\right]. \tag{56}$$

The log-likelihood function $\log p(\mathbf{r}_{p,t}|\underline{\theta})$ can be expressed as

$$\log p(\mathbf{y}_{p,t}|\underline{\theta}) = -\frac{1}{\sigma_n^2}\left\|\mathbf{y}_{p,t} - \alpha_{p,t,1}\mathbf{F}_t^T\mathbf{a}^{(b)}\left(\theta_{p,t,1}^{(b)}\right)^*\right\|^2 + \ln\frac{1}{\pi\sigma_n^2}. \tag{57}$$

The second derivative of $\log p(\mathbf{y}_{p,t}|\underline{\theta})$ can be obtained as

$$\begin{aligned}\frac{\partial^2 \log p(\mathbf{y}_{p,t}|\underline{\theta})}{\partial(\theta_{p,t,1}^{(b)})^2} =& -\frac{2}{\sigma_n^2}\alpha_{p,t,1}^*\frac{\partial(\mathbf{a}^{(b)}(\theta_{p,t,1}^{(b)}))}{(\partial\theta_{p,t,1}^{(b)})}(\theta_{p,t,1}^{(b)})\mathbf{F}_t^*\mathbf{F}_t^T\frac{\partial(\mathbf{a}^{(b)}(\theta_{p,t,1}^{(b)})^*)}{\partial\theta_{p,t,1}^{(b)}}\alpha_{p,t,1}\\ &-\frac{1}{\sigma_n^2}\left[\alpha_{p,t,1}^*\frac{\partial^2(\mathbf{a}^{(b)}(\theta_{p,t,1}^{(b)}))}{\partial(\theta_{p,t,1}^{(b)})^2}\mathbf{F}_t^*\mathbf{F}_t^T\mathbf{a}^{(b)}(\theta_{p,t,1}^{(b)})^*\alpha_{p,t,1}\right.\\ &+\alpha_{p,1,t}^*\mathbf{a}_b(\theta_{p,t,1}^{(b)})\mathbf{F}_t^*\mathbf{F}_t^T\frac{\partial^2(\mathbf{a}^{(b)}(\theta_{p,t,1}^{(b)})^*)}{\partial(\theta_{p,t,1}^b)^2}\alpha_{p,t,1}\\ &\left.+\mathbf{y}_{p,t}^H\mathbf{F}_t^T\frac{\partial^2(\mathbf{a}^{(b)}(\theta_{p,t,1}^{(b)})^*)}{\partial(\theta_{p,t,1}^{(b)})^2}\alpha_{p,t,1} +\alpha_{p,t,1}^*\frac{\partial^2(\mathbf{a}^{(b)}(\theta_{p,t,1}^{(b)}))}{\partial(\theta_{p,t,1}^{(b)})^2}\mathbf{F}_t^*\mathbf{y}_{p,t}\right].\end{aligned} \tag{58}$$

Using $E[\mathbf{y}_{p,t}] = \mathbf{F}_t^T\mathbf{a}^{(b)}(\theta_{p,t,1}^b)^*\alpha_{p,t,1}$, we can show that

$$(\mathbf{V})_{3,3} = \frac{2}{\sigma_n^2}\left\|\alpha_{p,t,1}\mathbf{F}_t^T\frac{\partial(\mathbf{a}^{(b)}(\theta_{p,t,1}^b)^*)}{\partial\theta_{p,t,1}^b}\right\|^2. \tag{59}$$

Next, $(\mathbf{V})_{1,3}$ and $(\mathbf{V})_{3,1}$ are expressed as

$$(\mathbf{V})_{3,1} = (\mathbf{V})_{1,3}^* = -E\left[\frac{\partial}{\partial\theta_{p,t,1}^{(b)}}\left(\frac{\partial\log p(\mathbf{y}_{p,t}|\underline{\theta})}{(\partial\alpha_{p,t,1}^{(b)})^*}\right)\right]. \tag{60}$$

One can easily show that

$$\begin{aligned}\frac{\partial}{\partial\theta_{p,t,1}^{(b)}}\left(\frac{\partial\log p(\mathbf{y}_{p,t}|\underline{\theta})}{(\partial\alpha_{p,t,1}^{(b)})^*}\right) =& -\frac{1}{\sigma^2}\left(\frac{\partial(\mathbf{a}^{(b)}(\theta_{p,t,1}^{(b)}))}{\partial\theta_{p,t,1}^{(b)}}\mathbf{F}_t^*\mathbf{F}_t^T\mathbf{a}^{(b)}(\theta_{p,t,1}^{(b)})^*\alpha_{p,t,1}\right.\\ &\left.+\mathbf{a}^{(b)}(\theta_{p,t,1}^{(b)})\mathbf{F}_t^*\mathbf{F}_t^T\frac{\partial(\mathbf{a}^{(b)}(\theta_{p,t,1}^{(b)})^*)}{\partial\theta_{p,t,1}^{(b)}}\alpha_{p,t,1} - \frac{\partial(\mathbf{a}^{(b)}(\theta_{p,t,1}^{(b)}))}{\partial\theta_{p,t,1}^{(b)}}\mathbf{F}_t^*\mathbf{y}_{p,t}\right).\end{aligned} \tag{61}$$

Similarly, using $E[\mathbf{y}_{p,t}] = \mathbf{F}_t^T\mathbf{a}^{(b)}(\theta_{p,t,1}^b)^*\alpha_{p,t,1}$, we can get

$$(\mathbf{V})_{3,1} = (\mathbf{V})_{1,3}^* = \frac{1}{\sigma^2}\alpha_{p,t,1}\frac{\partial\mathbf{a}^{(b)}(\theta_{p,t,1}^{(b)})}{\partial\theta_{p,t,1}^{(b)}}\mathbf{F}_t^*\mathbf{F}_t^T\mathbf{a}^{(b)}(\theta_{p,t,1}^{(b)})^*. \tag{62}$$

Finally, $(\mathbf{V})_{1,1}$ can be obtained as

$$(\mathbf{V})_{1,1} = -E\left[\frac{\partial}{\partial \alpha_{p,t,1}}\left(\frac{\partial \log p(\mathbf{y}_{p,t}|\underline{\theta})}{(\partial \alpha_{p,t,1})^*}\right)\right] \tag{63}$$

$$= E\left[\frac{1}{\sigma^2}\left(\mathbf{a}^{(b)}(\theta_{p,t,1}^{(b)})\mathbf{F}_t^*\mathbf{F}_t^T\mathbf{a}^{(b)}(\theta_{p,t,1}^{(b)})^*\right)\right] \tag{64}$$

$$= \frac{1}{\sigma^2}\left\|\mathbf{F}_t^T\mathbf{a}^{(b)}(\theta_{p,t,1}^{(b)})^*\right\|^2. \tag{65}$$

REFERENCES

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

[2] Y. Niu, Y. Li, D. Jin, L. Su, and A. V. Vasilakos, "A survey of millimeter wave (mmWave) communications for 5G: Opportunities and challenges," *Wireless Networks*, vol. 21, no. 8, pp. 2657-2676, Nov. 2015.

[3] R. W. Heath Jr, N. G. Prelcic, S. Rangan, W. Roh, and A. Sayeed, "An overview of signal processing techniques for millimeter wave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 435-453, April 2016.

[4] H. Xu, V. Kukshya, and T. S. Rappaport, "Spatial and temporal characteristics of 60GHz indoor channel," *IEEE J. Sel. Areas Commun.*, vol. 20, no. 3, pp. 620-630, April 2002.

[5] R. Daniels and R. Heath, "60GHz wireless communications: emerging requirements and design recommendation," *IEEE Veh. Technol. Mag.*, vol. 45, no. 12, pp. 41-50, Sept. 2007.

[6] T. S. Rappaport, R. W. Heath Jr., R. C. Daniels, and J. Murdock, Millimeter Wave Wireless Communications. Prentice-Hall, Sept. 2014.

[7] W. Roh, J. Seol, J. Park, B. Lee, J. Lee, Y. Kim, J. Cho, K. Cheun, and F. Aryanfar, "Millimeter-wave beamforming as an enabling technology for 5G cellular communications: theoretical feasibility and prototype results," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 106-113, Feb. 2014.

[8] S. Kutty and D. Sen, "Beamforming for millimeter wave communications: An inclusive survey," *IEEE Commun. Surveys and Tutorials*, vol. 18, no. 2, pp. 949-973, Secondquarter 2016.

[9] 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. Topic Signal Process.*, vol. 8, no. 5, pp. 831-846, Oct. 2014.

[10] A. Alkhateeb, O. E. Ayach, G. Leus, and R. W. Heath Jr, "Hybrid precoding for millimeter wave cellular systems with partial channel knowledge," *Proc. Inf. Theory and Appl. Workshop*, Feb. 2013.

[11] D. D. Donno, J. Beltran, D. Giustiniano, and J. Widmer, "Hybrid analog-digital beam training for mmWave systems with low-resolution RF phase shifters," *Proc. IEEE Inter. Conf. Commun. (ICC)*, May 2016.

[12] V. Venkateswaran and A. van der Veen, "Analog beamforming in MIMO communications with phase shift networks and online channel estimation," *IEEE Trans. Signal Process.*, vol. 58, no. 8, pp. 4131-4143, Aug. 2010.

[13] J. Wang *et al.*, "Beam codebook based beamforming protocol for multi-Gbps millimeter-wave WPAN systems," *IEEE J. Sel. Areas Commun.*, vol. 27, no. 8, pp. 1390-1399, Aug. 2009.

[14] A. Alkhateeb, M. Janhua, N. Gonzalez-Prelcic, and R. W. Heath Jr., "MIMO precoding and combining solutions for millimeter-wave systems," *IEEE Commun. Mag.*, vol. 52, no. 12, pp. 122-131, Dec. 2014.

[15] O. E. Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, Jr, "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1499-1513, Mar. 2013.

[16] B. Li, Z. Zhou, H. Zhang, and A. Nallanathan, "Efficient beamforming traiing for 60-GHz millimeter-wave communications: a novel numerical optimization framework," *IEEE Trans. Veh. Technol.*, vol. 63, no. 2, pp. 703-717, Feb. 2014.

[17] W. Yuan, S. M. D. Armour, and A. Doufexi, "An efficient and low-complexity beam training technique for mmWave communication," *Proc. IEEE Personal, Indoor, and Mobile Radio Commun. (PIMRC) 2015*, Sept. 2015.

[18] Y. Tsang, A. Poon, and S. Addepalli, "Coding the beams: Improving beamforming training in mmwave communication systems," in *Proc. IEEE Global Telecomm. Conf. (GLOBECOM)*, 2011, pp. 1-6.

[19] J. Wang, Z. Lan, W. W. Pyo, T. Baykas, C. S. Sum, M. Arizur Rahman, R. Funada, F. Kojima, I. Lakkis, H. Harada, and S. Kato, "Beam codebook based beamforming protocal for multi-Gbps millimeter-wave WPAN systems," *IEEE J. Selected Areas of Commun.,* vol. 27, no. 8, pp. 1390-1399, Oct. 2009.

[20] J. He, T. Kim, H. Ghauch, K. Liu, and G. Wang, "Millimeter wave MIMO channel tracking systems," *Proc. IEEE IEEE Global Telecomm. Conf. (GLOBECOM) Workshops*, Dec. 2014, pp. 416-421.

[21] W. Va, H. Vikalo, and R. W. Heath, "Beam tracking for mobile millimeter wave communication systems," *Proc. IEEE Global Conf. Signal and Inform. Process. (GlobalSIP)*, Dec. 2016, pp. 1-6.

[22] J. Brady, N. Behdad, and A. M. Sayeed, "Beamspace MIMO for millimeter-wave communications: System architecture, modeling, analysis and measurements," *IEEE Trans. Antennas Propag*., vol. 61, no. 7, pp. 3814-3827, July 2013.

[23] J. Palacios, D. D. Donno, and J. Widmer, "Tracking mm-Wave channel dynamics: fast beam training strategies under mobility," *Proc. IEEE Inter. Conf. Computer Commun. (Infocom) 2017*,

[24] 3GPP TS 36.211 V12.0.0 (2013-12): "Evolved Universal Terrestrial Radio Access (E-UTRA); Physical Channels and Modulation (Release 12)"

[25] A. van den Bos, "A cramer-rao lower bound for complex parameters," *IEEE Trans. on Signal Process.*, vol. 42, no. 10, pp. 2859-2859, Oct. 1994.

[26] J. W. Choi, B. Shim, Y. Ding, B. Rao, and D. I. Kim, "Compressed sensing for wireless communications: useful tips and tricks." *IEEE Commun. Surveys and Tutorials*, vol. 19, no. 3, pp. 1527-1550, Thirdquater 2017.

[27] J. A. Tropp, and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. on Inform. Theory.*, vol. 53, no. 12, pp. 4655-4666, Dec. 2007.

[28] H. V. Poor, *An introduction to signal detection and estimation, Second edition*, Springer.

[29] Z. Yang and X. Wang, "Joint mobility tracking and hard hand off in cellular networks via sequential monte carlo filtering," *Proc. IEEE Inter. Conf. Computer Commun. (Infocom) 2002*.

[30] C. Zhang, D. Guo, and P. Fan "Tracking angles of departure and arrival in a mobile millimeter wave channel," *Proc. IEEE Inter. Conf. Commun. (ICC)*, May 2016.

[31] J. Zhao, F. Gao, W. Jia, S. Zhang, S. Jin, and H. Lin "Angle domain hybrid precoding and channel tracking for mmWave massive MIMO systems," *IEEE Trans. Wireless Commun*., vol. 16, no. 10, pp. 6868-6880, Oct. 2017.