Dedicated Beam-based Channel Training Technique for Millimeter Wave Communications with High Mobility

Jisu Bae, Sun Hong Lim, Jin Hyeok Yoo and Jun Won Choi
Hanyang University, Seoul, Korea
Email: {jsbae, shlim, jhyoo}@spo.hanyang.ac.kr and junwchoi@hanyang.ac.kr

Byonghyo Shim
Seoul National University, Seoul, Korea
Email: bshim@islab.snu.ac.kr

***Abstract*—In this paper, we propose a new beam training framework to cope with mobility scenarios in millimeter wave communications. When a position of the mobile changes, the base-station needs to perform beam training frequently to track the time-varying channel, which leads to significant training overhead in radio resources. In order to alleviate this problem, we propose a "dedicated beam training" which serves only users under high mobility. Combined with conventional common beam training, the proposed dedicated beam training can allow the high mobility users to acquire channels with a small number of training beams exploiting the location information of the target user. The optimal selection of the training beams is formulated such that the lower bound of the angle of departure (AoD) estimate is minimized over the beam codebook indices given the estimate of the previous AoD state. Our numerical evaluation demonstrates that the proposed beam training scheme can maintain good channel estimation performance with less training overhead than the conventional beam training protocol.**

I. Introduction

The next generation wireless communication systems aim to achieve 1000 fold increase in throughput over the current 4G LTE systems [1], [2]. As a means to achieve this relentless goals, wireless communications using millimeter wave (mmWave) band has received much attention in recent years. One well-known drawback of mmWave communications is the significant attenuation of the signal power due to high free space path loss. In particular, the overall path loss is severe when the signal goes through rain and foliage or is blocked by obstacles, building, and human body. Recently, various studies have been conducted to overcome these limitations [1], [2]. The key enabler in these studies is the high directional beamforming using a large number of antennas to direct the signal power in the desired direction.

In order to support high directional beamforming, the base-station should acquire the channel state information (CSI) in the downlink. The *beam training* protocol for acquiring the CSI is as follows. First, the base-station sends the training beams one at a time to the different direction. Then, each user estimates its own CSI which is the composite of the direction of departure (DoA), the direction of arrival (AoA), and the channel gain, using the received signals [3]–[5]. The acquired CSI is then sent back to the base-station and used in computing the beamforming (precoding) matrix for the subsequent transmission of data symbols [6].

In the widely used beam training method so called *beam cycling*, the base-station transmits N training beams at the equally spaced directions. In order to support all users in all possible locations, the base-station should transmit a large number of beams. In this type of training referred to as *common beam training*, the training beams are shared by all scheduled users so that the training period should be sufficiently large to acquire reliable CSI for all users. This issue becomes more serious concern in mobility scenario where the location of mobiles is changing. In this case, clearly, beam training should be performed more frequently for the reliable channel tracking [4]. Note that the frequent use of common beam training is not a cost-efficient way since the small number of high mobility users in general very small.

An aim of this paper is to present a new beam training technique to support mobility scenarios in mmWave communications. The key ingredient in the proposed scheme is the *dedicated beam training* to serve high mobility users. When a user of high mobility is detected, the base-station transmits the dedicated beams to the user. Since the dedicated beams are intended for a particular user, the base-station can exploit the information on the user's location in selecting the beams. Specifically, using the knowledge on the AoD reported by the user, the base-station selects beam indices from the beam codebook that achieve the best channel estimation performance. Towards this end, minimizing (the lower bound of) variance of the estimation error on AoD over possible beam combinations. By employing a small number of dedicated beams and using a small number of resources allocated to the user, significant reduction in the training overhead can be achieved while maintaining good channel estimation performance.

It is worth comparing our approach with the previous beam tracking approaches [7]–[11]. In the previous studies, the AoD estimate is obtained from the channel tracking algorithm and then, a single beam is placed in the direction of the user. Since the user cannot estimate the channel accurately using the single beam-based measurement so that the seamless tracking is not possible when the angle of beam deviates from the true AoD. In

ISBN 978-90-827970-1-5

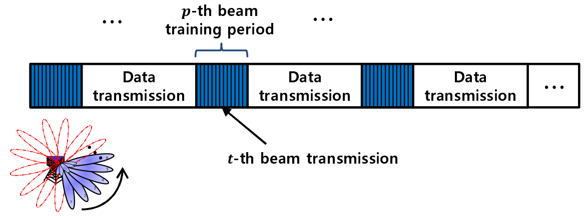

Fig. 1. Structure of common beam training

contrast, our scheme searches for the optimal placement of multiple beams generating the best channel estimation performance so that more robust and stable channel tracking is possible. We will show this in simulations section.

II. System Model

In this section, we discuss the system model for mmWave communications. We also discuss the procedure for conventional common beam training. In the conventional beam cycling scheme, the base-station transmits the beams one at a time at T equally partitioned directions between $[-\pi/2, \pi/2]$. Transmission of T beams is repeated every T_c seconds (see Fig. 1).

A. Channel Model

Consider a mmWave MIMO system with N_b base-station antennas and N_m user antennas. Adopting the angular-domain channel representation, the multi-input multi-output (MIMO) channel from the base-station to the ith user can be expressed as [5]

$$\mathbf{H}_{i,p} = \sum_{l=1}^{L} \alpha_{i,p,l} \mathbf{a}_m^{(i)}(\theta_{i,p,l}^m) \mathbf{a}_b^H(\theta_{i,p,l}^b) \tag{1}$$

where L is the number of multi-paths, p is the pth beam training period, $\alpha_{i,p,l}$ is the l-th path gain and $\theta_{i,p,l}^b$ and $\theta_{i,p,l}^m$ are the l-th path AoD and AoA, respectively. We assume that the channel $\mathbf{H}_{i,p}$ does not change during the beam training period. The beam steering vectors $\mathbf{a}_b(\theta_l^b)$ and $\mathbf{a}_m(\theta_l^m)$ are given by [5]

$$\mathbf{a}_b(\theta) = \frac{1}{\sqrt{N_b}}[1, e^{\frac{j2\pi d\theta}{\lambda}}, e^{\frac{j2\pi 2d\theta}{\lambda}}, \cdots, e^{\frac{j2\pi(N_b-1)d\theta}{\lambda}}]^T$$

$$\mathbf{a}_m(\theta) = \frac{1}{\sqrt{N_m}}[1, e^{\frac{j2\pi d\theta}{\lambda}}, e^{\frac{j2\pi 2d\theta}{\lambda}}, \cdots, e^{\frac{j2\pi(N_m-1)d\theta}{\lambda}}]^T$$

where d is the distance between the adjacent antennas and λ is the wavelength. Note that $\theta = \sin(\phi)$ where $\phi \in [-\frac{\pi}{2}, \frac{\pi}{2}]$ is a physical angle for AoD and AoA.

B. Conventional Common Beam Training

The signal received by the ith user for the tth beam transmission and the pth beam training period is given by

$$\begin{aligned} \mathbf{y}_{i,p,t} &= \mathbf{W}_{i,p}^H \mathbf{H}_{i,p} \mathbf{f}_t s_{i,t} + \mathbf{W}_{i,p}^H \mathbf{n}_{i,p,t} & (2) \\ &= \mathbf{W}_{i,p}^H \mathbf{H}_{i,p} \mathbf{f}_t s_{i,t} + \mathbf{n}'_{i,p,t} & (3) \end{aligned}$$

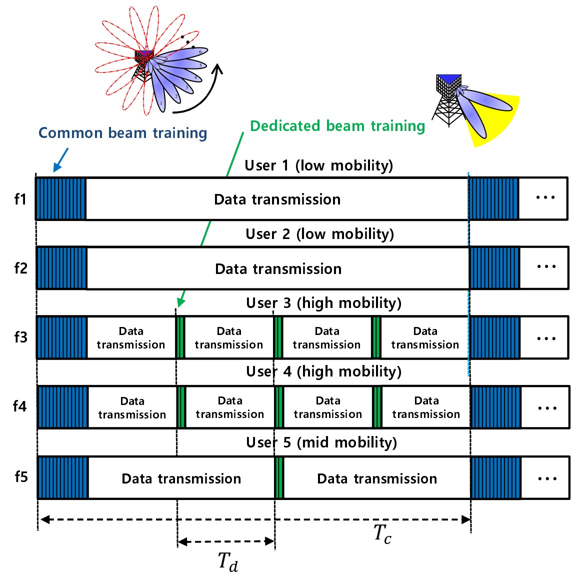

Fig. 2. An illustration of the proposed dedicated beam training.

where $\mathbf{W}_{i,p}^H$ is the combiner operation applied by the ith user, $\mathbf{f}_t$ is the beamforming vector used for the tth beam transmission, $\mathbf{n}_{i,p,t}$ is the Gaussian noise vector $\sim N(0, \sigma_n^2 \mathbf{I})$. Since $s_{i,t}$ is the known symbol, we simply set $s_{i,t} = 1$. The conventional beam cycling uses the beamforming vectors given by $\mathbf{f}_t = \mathbf{a}^{(b)}(-1 + 2(t-1)/T)$, where $t = 1, ..., T$. Collecting the measurements corresponding to T beam transmissions, we obtain the measurement matrix given by

$$\mathbf{Y}_{i,p} = \mathbf{W}_{i,p}^H \mathbf{H}_{i,p} \mathbf{F} + \mathbf{N}_{i,p}, \tag{4}$$

where $\mathbf{Y}_{i,p} = [\mathbf{y}_{i,p,1}, ..., \mathbf{y}_{i,p,T}]$, $\mathbf{F} = [\mathbf{f}_1, ..., \mathbf{f}_T]$, and $\mathbf{N}_{t,p} = [\mathbf{n}'_{i,p,1}, ..., \mathbf{n}'_{i,p,T}]$. Using the angular channel representation in (1), we have

$$\mathbf{Y}_{i,p} = \sum_{l=1}^{L} \alpha_{i,p,l} \mathbf{W}_{i,p}^H \mathbf{a}_m^{(i)}(\theta_{i,p,l}^m) \mathbf{a}_b^H(\theta_{i,p,l}^b) \mathbf{F} + \mathbf{N}_{i,p}. \tag{5}$$

The user acquires the channel matrix $\mathbf{H}_{i,p}$ by jointly estimating the channel gains $(\alpha_{i,p,1}, ..., \alpha_{i,p,L})$, the AoDs $(\theta_{i,p,1}^b, ..., \theta_{i,p,L}^b)$, and the AoAs $(\theta_{i,p,1}^m, ..., \theta_{i,p,L}^m)$. Since joint estimation of the parameters is computationally demanding, an approach exploiting the sparse nature of mmWave channels has been popularly used to reduce the computational complexity of the parameter estimation [12].

III. Proposed Beam Training for Mobility Scenario

Fig. 2 depicts the example of the proposed beam training protocol. We assume that each user is occupying different radio resources. First, the base-station monitors the extent of mobility for all scheduled users and detects high mobility users (user 3, 4, and 5 in the example). If the high mobility user is detected, then, the dedicated beam training mode is activated and the dedicated beams are transmitted to the designated users

ISBN 978-90-827970-1-5 © EURASIP 2018

between the common beam periods. As illustrated in Fig. 2, relatively small number of beam transmissions are enough for this purpose. The dedicated team transmission is repeated every T_d seconds. In order to track the time-varying channel, the frequency of dedicated beam training should be higher than that of common beam training (i.e., $T_c > T_d$). Based on the received signals, the user estimates the CSI and feeds it back to the base-station. Given the estimate of previous CSI, the base-station searches for the best dedicated beams for the next beam training period.

A. Dynamic mmWave Channel Models

As mentioned, channel estimation process consists of the estimation of the AoD, AoA, and channel gain. These channel parameters are time-varying when the users are moving in mobility scenario. The temporal behavior of $\theta^b_{i,p,l}$ and $\theta^m_{i,p,l}$ can be modeled by the discrete-time Markov random process described by the conditional probability of the current AoD/AoA state given the previous AoD/AoA state. One example of the conditional probability $Pr(\theta^b_{i,p,l}|\theta^b_{i,p-1,l})$ for AoD is Gaussian distribution, i.e., $N(\theta^b_{i,p-1,l}, \sigma^2_{AoD})$. Note that the parameter σ^2_{AoD} is proportional to mobility of the user. Hence, in practical systems, we can empirically map the average speed of the user to the appropriate value of σ_{AoD}. While the AoA can be modeled similarly, the temporal variation of channel gain $\alpha_{i,p,l}$ can be described by the autoregressive (AR) model, that is, $\alpha_{i,p,l} = a\alpha_{i,p-1,l} + \sigma_a\sqrt{1-a^2}u_p$, where a is the AR parameter, σ^2_a is the variance of the AR process, and u_p is the normal Gaussian process.

B. Beam Transmission and Channel Estimation

In the dedicated beam training, the base-station finds the beamforming vectors optimized for each user. Let $\mathbf{F}_{i,p} = [\mathbf{f}_{i,p,1}, ..., \mathbf{f}_{i,p,T}]$ be the beamforming matrix for the ith user and pth beam training cycle, then the base-station finds the optimal beamforming matrix $\mathbf{F}^*_{i,p}$ accounting for the conditional channel distribution given the previous CSI reported by the user. Since the beamforming matrix is optimized for each user, the required number of beam transmissions T is small. When the optimized beamforming matrix $\mathbf{F}^*_{i,p}$ is used, the measurement matrix in (5) can be expressed as

$$\mathbf{Y}_{i,p} = \sum_{l=1}^{L} \alpha_{i,p,l}\mathbf{W}^H_{i,p}\mathbf{a}^{(i)}_m(\theta^m_{i,p,l})\mathbf{a}^H_b(\theta^b_{i,p,l})\mathbf{F}^*_{i,p} + \mathbf{N}_{i,p}. \quad (6)$$

Note that the channel parameters $\{\alpha_{i,p,1}, ..., \alpha_{i,p,L}\}$, $\{\theta^b_{i,p,1}, ..., \theta^b_{i,p,L}\}$, and $\{\theta^m_{i,p,1}, ..., \theta^m_{i,p,L}\}$ are estimated by applying the Bayesian filter (e.g. extended Kalman filter [9]) using the sequence of observations $\{\mathbf{Y}_{i,1}, ..., \mathbf{Y}_{i,p}\}$. We can also estimate the CSI independently for each received vector using the compressed sensing recovery algorithms [12].

C. Beam Selection

In this subsection, we discuss the optimal beam placement for the dedicated beam training. Our goal is to choose the best beamforming vectors from the beam codebook $\mathcal{D}$ maximizing the channel estimation performance of the target user. The beam codebook $\mathcal{D}$ enumerates a variety of beamforming vectors with different beam-widths and with different steering directions. We first consider the beam design for single path scenario ($L = 1$) for simplicity but our result can be readily extended to the multi-path scenario by employing all beams optimized for each path. In the single path scenario, the received vector can be represented by

$$\mathbf{Y}^{(p)}_i = (\mathbf{W}^{(p)}_i)^H\mathbf{a}_m(\theta^m_{i,p,1})\alpha_{i,p,1}\mathbf{a}^H_b(\theta^b_{i,p,1})\mathbf{F}_{i,p} + \mathbf{N}^{(p)}_i. \quad (7)$$

Channel estimation performance can be quantified by analyzing the estimation errors for $\theta^b_{i,p,1}$, $\theta^m_{i,p,1}$, and $\alpha_{i,p,1}$. Since the impact of AoA information on the beam selection is insignificant, we assume that the AoA information is perfectly known by the base-station. Then, by letting $\mathbf{W}^{(p)}_i = \mathbf{a}_m(\theta^m_{i,p,1})$ in (7), we have

$$\mathbf{Y}^{(p)}_i = \alpha_{i,p,1}\mathbf{a}^H_b(\theta^b_{i,p,1})\mathbf{F}_{i,p} + \mathbf{N}^{(p)}_i. \quad (8)$$

In order to find the performance metric for beam selection, we derive the lower bound of the variance of the channel estimation error. First, the Cramer Rao Lower Bound (CRLB) for joint estimation of $\theta^b_{i,p,1}$ and $\alpha_{i,p,1}$ is given by

$$CRLB(\xi) > I^{-1} \quad (9)$$

$$I_{i,j} = -E\left[\frac{\partial^2 \ln P(\mathbf{Y}^{(p)}_i|\theta^b_{i,p,1}, \alpha_{i,p,1})}{\partial\xi^*_i\partial\xi_j}\right] \quad (10)$$

where $\xi = [\alpha^*_{i,p,1}, \alpha_{i,p,1}, \theta^b_{i,p,1}]^T$ and I is the Fisher information matrix. Once the estimate of $\theta^b_{i,p,1}$ is obtained, we can estimate the channel gain by projecting $\mathbf{Y}^{(p)}_i$ onto the space spanned by $\mathbf{a}^H_b(\theta^b_{i,p,1})\mathbf{F}_{i,p}$. Thus, it suffices to consider the CRLB for $\theta^b_{i,p,1}$ [13]

$$CRLB(\theta^b_{i,p,1}) = [I^{-1}]_{3,3} \quad (11)$$

$$= [Q - 2Re\{PCP^H\}]^{-1} \quad (12)$$

where $Q = \frac{1}{\sigma^2}\left\|\alpha_{i,p,1}\mathbf{F}^T_{i,p}\frac{\partial(\mathbf{a}^*_b(\theta^b_{i,p,1}))}{\partial\theta^b}\right\|^2$,
$P = \frac{1}{2\sigma^2}\alpha_{i,p,1}\mathbf{F}^T_{i,p}\frac{\partial\mathbf{a}_b(\theta^b_{i,p,1})}{\partial\theta^b}\mathbf{a}^*_b(\theta^b_{i,p,1})$, and
$C = \left(\frac{1}{2\sigma^2}\left\|\alpha_{i,p,1}\mathbf{F}^T_{i,p}\mathbf{a}^*_b(\theta^b_{i,p,1})\right\|^2\right)^{-1}$. Since the CRLB is the lower bound for the estimator of deterministic parameters, we average CRLB over the distribution of $\theta^b_{i,p,1}$. The distribution of the current AoD state given the previous AoD state can be obtained from the Baysian filter. Assuming that the distribution of $\theta^b_{i,p,1}$ follows Gaussian distribution with the mean $\hat{\theta}^b_{i,p,1}$ and the variance $\hat{\sigma}^2_{AoD}$, the average CRLB is

$$AvgCRLB = \int CRLB_{k,l}(\theta) \cdot N(\hat{\theta}^b_{i,p-1,1}, \hat{\sigma}^2_{AoD})d\theta.$$

The final step of the beam selection is to find out the index minimizing $AvgCRLB$ over all T combinations of beam indices from $\mathcal{D}$. Although this this process requires to high

ISBN 978-90-827970-1-5 © EURASIP 2018

search complexity in general, when we use only small value of T, the complexity can be made resonally small. In particular, when we use dual beams (i.e., $T = 2$), we can search for the beam indices for $\mathbf{f}_{i,p,1}$ and $\mathbf{f}_{i,p,2}$ over two dimensional grid. The search complexity can be further reduced by using the fact that the optimal beamforming pair is mostly symmetric with each other with respect to the previous AoD estimate $\hat{\theta}^b(t-1)$. Hence, we can perform one dimensional grid search and apply the look-up table for mapping the previous AoD estimate and variance to the beam indices.

D. Proposed Beam Tracking for Multi Path Scenarios

So far, we have presented the new beam tracking strategy for single path scenarios. We can easily extend the proposed scheme for the scenario where there exist L multi paths in mmWave channels. If the AoDs associated with each path are well separated in angular domain, it is possible to apply the proposed tracking scheme derived for single path for each individual path while ignoring the existence of other paths. In this scenario, the base-station transmits two training beams for each of L path, requiring $2L$ beam transmissions in total. Since we search for the AoD estimate within the restricted range, we can separate each path from each other without negligible performance loss. When the different paths are clustered in angular domain, we have to find joint estimate of AoDs based on the received signals generated from $2L$ beam transmissions. The optimization for designing $2L$ beamforming vectors can be performed for each path. The estimation of L values of AoD can be performed via compressed sensing techniques such as orthogonal matching pursuit (OMP) [14].

IV. Simulation Results

In this section, we evaluate the effectiveness of the proposed beam training method. We consider the base-station and the mobile equipped with $N_b = N_m = 32$ antennas. We assume uniform linear arrays (ULAs). We use 32 beams for the common beam training. OMP algorithm is used to estimate the AoA and AoD, which are generated based on the statistical model we described in Subsection III-A. The range of AoD between $[-\pi/2, \pi/2]$ is discretized into 128 angular bins, respectively. On the other hand, we use the dual beams ($N = 2$) for the dedicated beam training. The beam codebook used for the proposed beam selection includes the steering vectors at 128 uniformly quantized directions. The OMP algorithm is also used for AoD estimation.

Fig. 3 shows the bit error rate (BER) performance as a function of training overhead, where the training overhead is the ratio of the number of symbols used for training to the total number of symbols in the frame. We generate 1000 frames to evaluate the performance of dedicated beam training. A frame contains 4000 symbols which include symbols for training and data transmission. We assume that the channel changes over symbol time. i.e., the channel of the nth symbol H_n is

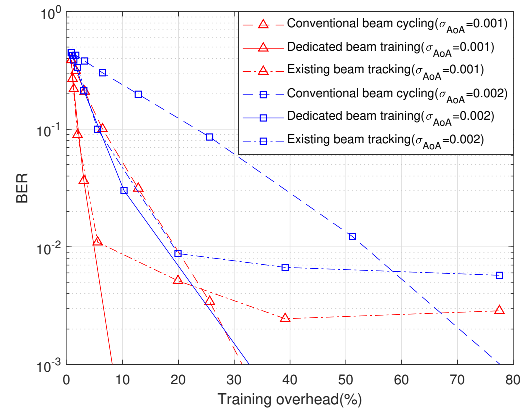

Fig. 3. BER versus training overhead.

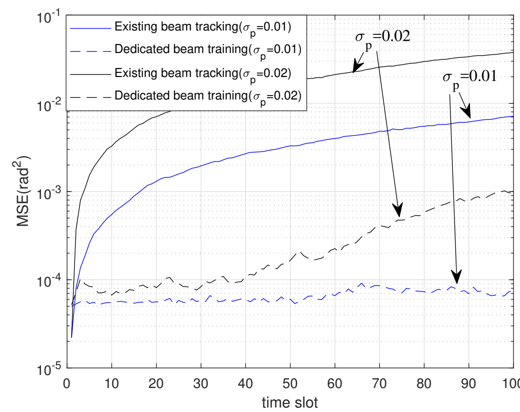

Fig. 4. MSE versus time slot for AoD estimation

generated by

$$H_n = \sum_{l=1}^{L} \alpha_{n,l} a_m(\theta^m_{n,l}) a_b(\theta^b_{n,l})^H$$

where nth channel states of lth path are follows channel gain $\alpha_{n,l} = \rho\alpha_{n-1,l} + v_t\sqrt{1-\rho^2}$, AoD $\theta^b_{n,l} \sim N(\theta^b_{n-1,l}, \sigma^2_{AoD})$ and AoA $\theta^m_{n,l} \sim N(\theta^m_{n-1,l}, \sigma^2_{AoA})$. We evaluate the performance of 1) the common beam training, 2) the dedicated beam training, and 3) existing beam tracking [9]. The existing beam tracking method exploits only one beam pair and estimates the AoD and AoA based on extended kalman filter (EKF). All of the methods transmit $N_t = 32$ beams in the beginning of the frame. While the beam cycling be repeated every T_c symbols, the dedicated beam training and the method in [9] are performed every T_d symbols. The remaining symbol slots are allocated for data symbols where binary phase shift keying (BPSK) modulation is used. The precoding matrix for the data transmission is the matrix V obtained from taking the singular value decomposition of estimated channel $\hat{H}_n = \sum_{l=1}^{L} \hat{\alpha}_{n,l} a_m(\hat{\theta}^m_{n,l}) a_b(\hat{\theta}^b_{n,l})^H$. We consider the scenario

ISBN 978-90-827970-1-5 © EURASIP 2018

with $L = 3$ and $SNR = 20$dB. We set $\rho = 0.9999$ and $\sigma_{AoD} = \sigma_{AoA} = 0.001$. Although the variation of the channel states looks very slow, since the channel states vary across hundreds symbols or more, these are changing quite fast. For example, if the beam training repeated every 400 symbols, the variation of channel states in the adjacent beam training corresponds to $\rho' \approx 0.96$ and $\sigma'_{AoD} = \sigma'_{AoA} = 0.02(\approx 1.14°)$.

We compare the training efficiency of the common beam training and the proposed beam training. With $T_c = 4000$, the training overhead of common beam training corresponds to $0.8\%(= N_t/400 * 100)$. As T_c decreases to $T_c = 400$ and 50, training overhead increase to 8% and 64%. In contrast, since the dedicated beam training only transmits $2L$ training symbols instead of N_t beams, the training overhead corresponds to only 2.1% and 12.8% with $T_d = 400$ and 50. Even if the existing tracking method with only one beam pair has lower training overhead with same T_d, the BER performance of existing beam tracking is saturated. This is because the estimate of the existing beam tracking quickly outdated. One can know this through simulation result in Fig. 4. Note that the training overhead of the proposed method would be much lower at same BER performance. Such substantial performance gain is attained since the proposed scheme uses the beam directions chosen to bring the optimal AoD estimation quality while the common beam training uses equally spaced beams without using the information on the previous AoD.

Fig.4 shows the tracking performance of proposed method and existing beam tracking method. We set the number of antennas $N_b = N_m = 32$, $SNR = 20$dB, and the standard deviation of AoA and AoD to $\sigma_p = \{0.01, 0.02\}$ ($\approx \{0.57°, 1.14°\}$). Even if the variation of the AoD and AoA seems too small, since the duration of a time slot is a few millisecond, the AoA and AoD actually change at high speed in this scenario. We repeated the 1000 simulations to compare the proposed method with existing beam tracking method [9]. In order to compare our scheme with existing beam tracking method, we select the beam pair for both user side and base-station side, and find the AoA and AoD using OMP. The proposed method shows much better tracking performance rather than the method in [9]. By using two beams to track the AoD, the dedicated beam training performs much more stable tracking performance. Since the existing beam tracking method exploit only one beam pair, while the estimation error keeps rapidly increasing, the proposed method has much better estimation accuracy.

V. Conclusions

In this paper, we have presented the novel beam technique suitable for the mobility scenario in mmWave communications. To this end, we proposed the dedicated beam training which transmits beams for a particular user. We demonstrated that the beam selection algorithm using the information on user's location and channel, the dedicated beam training can reduce the training overhead significantly while maintaining good BER performance.

VI. Acknowledgement

This work was supported by the National Research Foundation of Korea Grant funded by the Korean Government (NRF-2017R1D1A1A09000602) and Samsung Research Funding & Incubation Center of Samsung Electronics under Project Number SRFC-IT-1601-09.

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, 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, 2015.

[3] A. M. Sayeed, "Deconstructing multiantenna fading channels," *IEEE Trans. Signal Process.*, vol. 1, pp. 156-166, Jun. 2007.

[4] 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. Topic Signal Process.*, vol. 10, no. 3, pp. 436-453, April 2016.

[5] 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.

[6] 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.

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

[8] K. Gao, M. Cai, D. Nie, B. Hochwald, J. N. Laneman, H. Huang, and K. Liu, "Beampattern-based tracking for millimeter wave communication systems," *Proc. of the IEEE GLOBECOM, 2016.*

[9] V. Va, H. Vikalo, and R. W. Heath, "Beam tracking for mobile millimeter wave communication systems," *Proc. of IEEE GlobalSIP, Dec. 2016.*

[10] C. Zhang, D. Guo and P. Fan, "Tracking angles of departure and arrival in a mobile millimeter wave channel," *2016 IEEE International Conference on Communications (ICC)*, Kuala Lumpur, 2016, pp. 1-6.

[11] S. Jayapakasam, X. X. Ma, J. W. Choi, and S. Kim, "Robust beam-tracking for mmWave mobile communications," *To appear in IEEE Communications Letters*, DOI 10.1109.

[12] 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 Tuts.*, vol. 19, no. 3, pp. 1527-1550, thrdquarter 2017.

[13] A. Bos, "A cramer-rao lower bound for complex parameters," *IEEE Trans. on Signal Process.*, vol. 42, no. 10, pp. 2859-2589, Oct. 1994.

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