Impact of Training on mmWave Multi-User MIMO Downlink

Gilwon Lee, Jungho So, and Youngchul Sung
School of Electrical Engineering
KAIST
Daejeon 305-701, South Korea
Email: {gwlee@, jhso@, and ysung@ee.}kaist.ac.kr

***Abstract*—The impact of training on the performance of millimeter wave multi-user multiple-input multiple-output downlink systems based on hybrid analog/digital beamforming is investigated in the regime of a large number of transmit antennas under the uniform random single-path channel model. In particular, the performance loss with respect to the number of training beams is quantified for general training-based data transmission schemes with several different training beam designs when full training is not applicable. Our analysis provides insights into effective training beam design and simulation result validates our analysis.**

Keywords*—*millimeter wave MU-MIMO downlink, impact of training, hybrid beamforming architecture

I. Introduction

The millimeter (mmWave) technology is one of the leading candidates for future wireless access networks [1]. The mmWave technology can use the large bandwidth available in the mmWave spectrum from 30 to 300 GHz to provide high data rates based on the large bandwidth. However, the propagation characteristics in the mmWave band are not so friendly; mmWave signals experience large path loss and blockage. Overcoming this large path loss and blockage in the mmWave band is a major challenge to realizing the mmWave technology. Fortunately, due to the small wavelengths of mmWave signals, large antenna arrays can be placed into small physical spaces and it is possible to perform highly directional beamforming based on large antenna arrays to compensate for the large path loss in the mmWave band. However, there are several practical challenges to highly directional beamforming based on large antenna arrays. The first is the hardware limitation. That is, if each antenna is attached to a radio-frequency (RF) chain and an analog-to-digital or digital-to-analog converter, the amount of required hardware is too large since there are many antenna elements in a large antenna array. The second is that there exist too many beamforming directions to initially search to find the desired users and their channel gains with a narrow transmit beam with a high beamforming gain. In other words, training overhead is large for large antenna arrays. There has been much effort to resolve the above issues both in single-user (SU) mmWave multiple-input multiple-output (MIMO) and multiuser (MU) mmWave MIMO, e.g., hybrid A/D beamforming and efficient training signal design [2]–[10]. In particular, in the case of mmWave MU-MIMO, which is the main focus of this paper, several hybrid A/D beamforming methods, beam selection or user selection methods are proposed [7]–[10]. In [7], a two-stage MU hybrid precoding method is proposed and its performance is analyzed in several regimes including the large-antenna-array regime and the limited feedback regime. In [8], a beam selection method is proposed for mmWave MU-MIMO systems with analog beamformers at the BS and each user. In [9] and [10], the performance of random beamforming is analyzed in sparse mmWave channels and new efficient scheduling methods are proposed by exploiting the sparsity of mmWave channels. However, most previous works on mmWave MU-MIMO systems including [7]–[10] implicitly assume full training which induces heavy training overhead.

In this paper, we consider mmWave MU-MIMO downlink systems based on hybrid A/D beamforming. Note that with hybrid A/D beamforming the channel estimation based on uplink signals and channel reciprocity is not directly applicable since the measured uplink channel at the baseband after digital-to-analog conversion is the product of the BS analog beamforming matrix and the actual wireless channel and the design of the BS analog beamforming matrix is left independent of the uplink signals. Thus, we here consider a general downlink training-based data transmission scheme (described by steps A.1 to A.3 in Section II) and investigate the impact of training on the performance of the general training-based mmWave MU-MIMO downlink system. In particular, we focus on the impact of the amount of training and the beam shape on the performance of mmWave MU-MIMO downlink and analyze the performance loss associated with the amount of training and the beam shape relative to full training in an asymptotic regime (for large scale antenna arrays) in which the number of transmit antennas tends to infinity.

II. System Model

We consider a mmWave MU-MIMO downlink system where the BS with M transmit antennas and M_{RF} $(\leq M)$ RF chains simultaneously serve K users with N receive antennas each. We assume that each user has a single RF chain and performs analog beamforming with the N receive antennas. Hence, the BS transmits only one stream to each user and the maximum number of users simultaneously served by the BS is the number of RF chains of the BS. So, we simply set $K = M_{RF}$. We assume that the BS performs hybrid beamforming to the K users with the overall beamforming matrix given by the product of an $M_{RF} \times K$ digital beamforming matrix $\mathbf{W} = [\mathbf{w}_1, \cdots, \mathbf{w}_K]$ at the baseband, where $\mathbf{w}_k$ is the digital beamforming vector for user k, and an $M \times M_{RF}$ analog beamforming matrix $\mathbf{U} = [\mathbf{u}_1, \cdots, \mathbf{u}_{M_{RF}}]$ at the RF band. Then, the precoded (beamformed) downlink signal $\mathbf{x}$ is given

This research was supported by the Cross-Ministry Giga KOREA Project of The Ministry of Science, ICT and Future Planning, Korea (GK14N0100, 5G mobile communication system development based on mmWave).

by $\mathbf{x} = \mathbf{U}\mathbf{W}\mathbf{d}$, where $\mathbf{d} = [d_1, \cdots, d_K]^T$ is the $K \times 1$ data symbol vector assumed to be a zero-mean Gaussian vector with variance P_k/K, i.e., $\mathbf{d} \sim \mathcal{CN}(\mathbf{0}, \frac{P_t}{K}\mathbf{I})$, and P_t is the average total transmit power. We normalize the digital beamforming matrix $\mathbf{W}$ such that $\|\mathbf{U}\mathbf{W}\|_F^2 = K$ to have the total transmit power constraint $P_t = \mathbb{E}[\|\mathbf{x}\|^2] = \frac{P_t}{K}\|\mathbf{U}\mathbf{W}\|_F^2$. Note that if $\mathbf{U}^H\mathbf{U} = \mathbf{I}$, the constraint $\|\mathbf{U}\mathbf{W}\|_F^2 = K$ is reduced to $\|\mathbf{W}\|_F^2 = K$. For simplicity, we consider a narrowband fading channel model. (For wideband channels, OFDM can be adopted and the considered model here corresponds to one subcarrier channel in this case.) Then, the received signal of user k at the N receive antennas is given by

$$\mathbf{r}_k = \mathbf{H}_k\mathbf{U}\sum_{m=1}^{K}\mathbf{w}_m d_m + \mathbf{n}_k, \tag{1}$$

where $\mathbf{H}_k$ is the $N \times M$ channel matrix that represents the downlink channel between the BS and user k, and $\mathbf{n}_k \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$ is the noise vector. With receive analog beamforming of user k, the final receiver output of user k is given by

$$y_k = \mathbf{v}_k^H\mathbf{H}_k\mathbf{U}\sum_{m=1}^{K}\mathbf{w}_m d_m + \mathbf{v}_k^H\mathbf{n}_k \tag{2}$$

where $\mathbf{v}_k$ is the analog beamforming vector of user k.

To model the sparsity of mmWave channels, we adopt the widely-used geometric channel model with L_k scatterers for user k and assume that the antenna configuration of the BS and each user k is the uniform linear array (ULA). Thus, the channel matrix $\mathbf{H}_k$ between the BS and user k is given by [11], [12]

$$\mathbf{H}_k = \sqrt{\frac{MN}{L_k}}\sum_{i=1}^{L_k}\alpha_{k,i}\mathbf{a}_R(\phi_{k,i})\mathbf{a}_T(\theta_{k,i})^H. \tag{3}$$

Here, $\alpha_{k,i}$ is the i-th path gain of user k; $\phi_{k,i}$ and $\theta_{k,i}$ are respectively the normalized angle-of-arrival (AoA) and angle-of-departure (AoD) of path i of user k generated according to the uniform distribution on the interval $[-1, 1)$, i.e., $\phi_{k,i}, \theta_{k,i} \overset{i.i.d.}{\sim} \text{Unif}[-1, 1]$; and the ULA steering and response vectors of the BS and user k are respectively given by $\mathbf{a}_T(\theta) = \frac{1}{\sqrt{M}}[1, e^{-\iota\pi\theta}, \cdots, e^{-\iota\pi(M-1)\theta}]^T$ and $\mathbf{a}_R(\phi) = \frac{1}{\sqrt{N}}[1, e^{-\iota\pi\phi}, \cdots, e^{-\iota\pi(N-1)\phi}]^T$, where $\iota = \sqrt{-1}$. *

A. The Considered Training-Based Data Transmission Scheme: Downlink Training, Feedback, and Achievable Rate

In this subsection, we describe the considered training-based data transmission scheme. In the considered scheme, we employ the classical MU-MIMO downlink training based on beam sweeping [13] adapted to hybrid A/D beamforming. The BS sequentially transmits analog training beam vectors to the space in the cell and each user estimates its channel by sweeping all of its receive analog beams for each transmit training beam vector. For example, in the full training method, the BS transmits all M orthogonal training beams available with M transmit antennas and each user uses all N orthogonal receive beams available with N receive antennas to estimate the $N \times M$ channel matrix. In this case, the total number of training overhead required for channel estimation is given by MN. However, it may be practically unreasonable to execute the full training method since a long training period is required for a large-scale antenna array at the BS operating in the mmWave band and this reduces the data rate significantly. Therefore, it is necessary that the BS performs partial training based on a smaller number S of training beams than the number M of transmit antennas, yielding a shorter training period. Then, a natural question arising in this case is "*what is the performance loss relative to the full training performance when only S ($K = M_{RF} \le S < M$) training beams are used and how should we design the training beam set $\{\mathbf{t}_1, \cdots, \mathbf{t}_S\}$?*" In this paper, we investigate this performance loss for several different training beam designs.

The overall procedure of the considered downlink training-based data transmission scheme is described in the below.

A.1 During the training period, the BS transmits S analog training beam vectors $\{\mathbf{t}_i\}_{i=1}^S$ to K users in the cell and each user k estimates its channel state information (CSI) by sweeping its N receive analog beam vectors $\{\mathbf{v}_{k,j}\}_{j=1}^N$ for each transmit training beam vector. Here, we assume that $\{\mathbf{t}_i\}_{i=1}^S$ and $\{\mathbf{v}_{k,j}\}_{j=1}^N$ are sequences of orthonormal vectors.

A.2 After the training period is over, each user k computes

$$\{\mathbf{v}_k^*, \mathbf{t}_k^*\} = \underset{\mathbf{v}_{k,j}, \mathbf{t}_i}{\arg\max}\ |\mathbf{v}_{k,j}^H\mathbf{H}_k\mathbf{t}_i|; \tag{4}$$

determines $\mathbf{v}_k^*$ as the receive analog beam vector to be used during the data transmission; and feeds some channel quality indicator (CQI) denoted by $\mathcal{F}_k$, which will be defined soon, back to the BS.

A.3 Based on $\mathcal{F}_k$, the BS properly designs the transmit analog beamforming matrix $\mathbf{U}$ and the digital beamforming matrix $\mathbf{W}$, and serves the K users with the designed beamforming matrices.

Several different feedback methods can be considered for the above scheme, and the achievable sum rate of the scheme depends on the type of CQI $\mathcal{F}_k$ from each user and is given as follows.

i) $\mathcal{F}_k$ = the corresponding beam index of $\mathbf{t}_k^*$ (partial CSI feedback): In this case, we can choose the analog beamforming matrix as $\mathbf{U}^* = [\mathbf{t}_1^*, \cdots, \mathbf{t}_K^*]$ and the digital beamforming matrix as $\mathbf{W} = \mathbf{I}$ to yield single-beam matched filtering. Then, the average achievable sum rate in this case is given by

$$\mathcal{R}_{sum}^{index} = \mathbb{E}\left[\sum_{k=1}^{K}\log\left(1 + \text{SINR}_k(\mathbf{U}^*, \mathbf{I})\right)\right] \tag{5}$$

where

$$\text{SINR}_k(\mathbf{U}, \mathbf{W}) = \frac{\frac{P_t}{K}|\bar{\mathbf{h}}_k^H\mathbf{U}\mathbf{w}_k|^2}{1 + \frac{P_t}{K}\sum_{m\neq k}|\bar{\mathbf{h}}_k^H\mathbf{U}\mathbf{w}_m|^2} \tag{6}$$

and $\bar{\mathbf{h}}_k^H := \mathbf{v}_k^{*H}\mathbf{H}_k$ is the receiver-filtered channel vector of user k. The amount of feedback required for this scheme is one integer for each user.

*The normalized AoA (AoD) $\theta \in [-1, 1]$ is related to the physical AoA (AoD) $\phi \in [-\pi/2, \pi/2]$ as $\theta = \frac{2d\sin(\phi)}{\lambda}$ [11], where d and λ are the distance between two adjacent antenna elements and the carrier wavelength, respectively. We assume the critical spatial sampling of $d/\lambda = 1/2$ in this paper.

ii) $\mathcal{F}_k = [\bar{\mathbf{h}}_k^H \mathbf{t}_1, \cdots, \bar{\mathbf{h}}_k^H \mathbf{t}_S]^T$ (full CSI feedback): In this case, we can design the analog and digital beamforming matrices at the BS by solving the following problem:

$$\mathcal{R}_{sum}^{full} = \mathbb{E}\left[\max_{\mathbf{U},\mathbf{W}} \sum_{k=1}^{K} \log\left(1 + \mathrm{SINR}_k(\mathbf{U},\mathbf{W})\right)\right], \quad (7)$$

under the conditions $\mathbf{U} = [\mathbf{t}_i]_{i\in\mathcal{I}}$ and $\|\mathbf{UW}\|_F^2 = K$, where $\mathcal{I}$ is the index set of chosen transmit beams out of the S training beams $\{\mathbf{t}_1, \cdots, \mathbf{t}_S\}$ with $|\mathcal{I}| = K = M_{RF} \leq S$. Note that the number of possible combinations for $\mathbf{U}$ is $\binom{S}{|\mathcal{I}|}$. The amount of feedback required for this scheme is $2S$ real numbers for each user.

iii) $\mathcal{F}_k$ = some CQI between *i)* and *ii)*: This case includes various feedback schemes. For example, the two-stage feedback scheme considered in [7] belongs to this class. The average sum rate of this class is different for different $\mathcal{F}_k$ but is lower and upper bounded as

$$\mathcal{R}_{sum}^{index} \leq \mathcal{R}_{sum}(\mathcal{F}_k) \leq \mathcal{R}_{sum}^{full}, \quad \forall\, \mathcal{F}_k. \quad (8)$$

B. The Considered Training Beam Vectors

In Section II-A, we did not specify the training beam set for beam sweeping. There can be many types of training beams [13]. For example, in the case of full training with M transmit antennas, one can consider $\mathbf{t}_i = [0, \cdots, 0, \underbrace{1}_{i-th}, 0, \cdots, 0]^T$, $\mathbf{t}_i = \mathbf{a}_T\left(\frac{2(i-1)}{M}\right)$, a randomly-spread beam over the entire AoD domain (typically encountered in the compressed sensing (CS) based channel estimation for mmWave MIMO), or the i-th column of any $M \times M$ unitary matrix. However, in this paper we consider the beamforming-type training beams since it is reasonable to assume that the training beam itself should be beamformed to identify channel elements by overcoming the large path loss in the mmWave band. To investigate the impact of width and coverage of partial beamforming-type training beams on the performance of mmWave MU-MIMO downlink, we consider the following three sets of S training beams which differ in the combination of width and angle coverage:

B.1 S consecutive narrow beams with the spacing of $\frac{2}{M}$ between two adjacent training beams in the normalized AoD domain:

$$\mathbf{t}_i^{(M)} = \mathbf{a}_T(\vartheta_i) = \mathbf{a}_T\left(\vartheta + \frac{2(i-1)}{M}\right), \quad (9)$$

for $i = 1, \cdots, S$, where $\vartheta \sim \mathrm{Unif}[-1, 1]$ is the random offset.

B.2 S narrow beams equally-spaced in the entire interval $[-1, 1]$ of the normalized AoD domain:

$$\mathbf{t}_i^{(S)} = \mathbf{a}_T\left(\vartheta + \frac{2(i-1)}{S}\right), \quad (10)$$

for $i = 1, \cdots, S$, where $\vartheta \sim \mathrm{Unif}[-1, -1 + \frac{2}{S}]$.

B.3 S wide beams each of which is the normalized sum of D consecutive narrow beams (Woodward Lawson synthesis) [14]:

$$\mathbf{u}_i^{(WB)} = \sqrt{\frac{1}{D}} \sum_{d=D(i-1)+1}^{Di} e^{-\iota\pi\vartheta_d(M-1)/2} \mathbf{a}_T(\vartheta_d), \quad (11)$$

for $i = 1, \cdots, S$, where $D = \frac{M}{S}$ and $\vartheta_d = \vartheta + \frac{2(i-1)}{M}$ with $\vartheta \sim \mathrm{Unif}[-1, -1 + 2/M]$, and we assume that S divides M for the sake of simplicity.

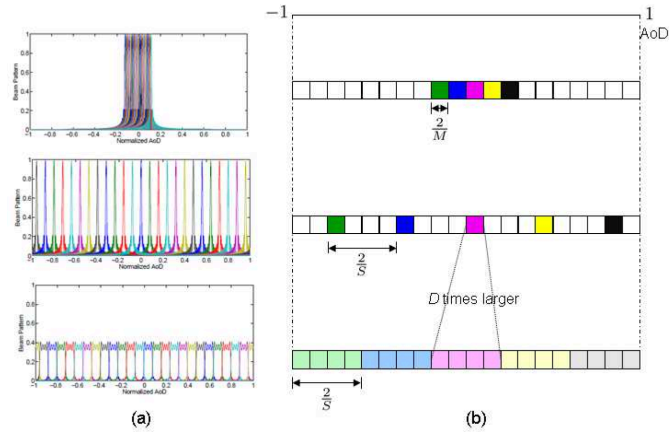

Fig. 1. (a) Beam patterns of the considered sets of S training beams (9), (10), and (11) versus normalized AoD (from top to bottom), and (b) the corresponding representations in the normalized AoD domain.

Note that the main lobe width of each narrow beam is given by $\frac{2}{M}$. In the case of the wide beam, the main lobe width is increased by factor D but the beam pattern gain is reduced by factor $\sqrt{D}$, as shown in Fig. 1(a). Fig. 1(b) shows the visual illustration of the three considered sets of training beams in the normalized AoD domain with sidelobes hidden.

For the receive analog beams, we assume that each user k uses a set of N orthonormal analog beamforming vectors given by

$$\mathbf{v}_{k,j} = \mathbf{a}_R\left(\varphi_k + \frac{2(j-1)}{N}\right), \quad \text{for} \quad j = 1, \cdots, N \quad (12)$$

where $\varphi_k \sim \mathrm{Unif}[-1, 1]$. In this paper, we assume $N \ll M$ and full receiver training with the N beam vectors at the receiver side.

III. Asymptotic Analysis of Impact of Training

Rate analysis of hybrid A/D beamforming systems is difficult since the constraints for analog and digital beamforming matrices are intertwined. Furthermore, mmWave MIMO channels are sparse in the AoD and AoA domains and thus many available results in rich scattering channel environments are not applicable. To circumvent these difficulties and analyze the impact of training on the rate performance of mmWave MU-MIMO downlink, we adopt the uniform random single-path (UR-SP) channel model capturing the highly-directional propagation in the mmWave band [2], [7], [9], [10], [15], [16], which is given by (3) with $L_k = 1$ for all k, and apply asymptotic techniques with the number M of transmit antennas tending to infinity to capture the situation of large-scale transmit antenna arrays. We set $K = M^q$ with $q \in (0, 1)$ and $S = M^\ell$ with $\ell \in [q, 1)$ for K and S as functions of M to implement our assumption $K \leq S < M$ in the asymptotic regime of M tending to infinity.

With the above assumptions, the main theorem regarding the rate performance of the three training beam designs B.1, B.2, and B.3 is given in the below.

Theorem 1: Under the UR-SP channel model (3) with $L_k = 1$ and fixed N and P_t, as $M \to \infty$ with $K = M^q$ ($q \in (0,1)$) and $S = M^\ell$ ($\ell \in [q,1)$), the average sum rate $\mathcal{R}_{sum}(\mathcal{F}_k)$ of the training-based data transmission scheme described by steps A.1, A.2 and A.3 scales as

$$\mathcal{R}_{sum}(\mathcal{F}_k) = \begin{cases} o\left(M^q\right) & \text{for B.1} \\ \Theta\left(M^q \log(1 + M^{2\ell-q-1})\right) & \text{for B.2} \\ \Theta\left(M^q \log(1 + M^{\ell-q})\right) & \text{for B.3,} \end{cases} \tag{13}$$

regardless of $\mathcal{F}_k$, where $o(\cdot)$ and $\Theta(\cdot)$ are the small-o and big-theta of Bachmann-Landau notation (asymptotic notation), respectively.

Proof: Proof is omitted due to space limitation. ■

It can be shown that the average sum rate $\mathcal{R}^\star_{sum}$ of the full training method scales as

$$\mathcal{R}^\star_{sum} = \Theta\left(K \log(1 + M/K)\right) \tag{14}$$

$$= \Theta(M^q \log(1 + M^{1-q})), \tag{15}$$

as $M \to \infty$. From Theorem 1, we have the following corollary.

Corollary 1: Eq. (13) in Theorem 1 can be re-expressed as

$$\mathcal{R}_{sum}(\mathcal{F}_k) = \begin{cases} o\left(K\right) & \text{for B.1} \\ \Theta\left(K \log(1 + \xi^2 \frac{M}{K})\right) & \text{for B.2} \\ \Theta\left(K \log(1 + \xi \frac{M}{K})\right) & \text{for B.3,} \end{cases} \tag{16}$$

where $\xi = \frac{S}{M}$ (< 1) is the ratio of the number of training beams to the number of transmit antennas.

Proof: (16) is simply obtained by substituting $M^q = K$ and $M^\ell = S$ into (13). ■

The relative rate performance of partial training to that of full training can be captured by the ratio μ of $\mathcal{R}_{sum}(\mathcal{F}_k)$ to $\mathcal{R}^\star_{sum}$, i.e., $\mu := \frac{\mathcal{R}_{sum}(\mathcal{F}_k)}{\mathcal{R}^\star_{sum}}$. Then, by combining the above results, we have the following theorem regarding μ:

Theorem 2: The ratio μ converges to

$$\mu \to \begin{cases} 0, & \text{for B.1 and } \ell \in (q,1) \\ \frac{2\ell-q-1}{1-q}, & \text{for B.2 and } \ell \in (\frac{1+q}{2},1) \\ \frac{\ell-q}{1-q}, & \text{for B.3 and } \ell \in (q,1) \end{cases} \tag{17}$$

as $M \to \infty$.

Proof: It is obtained by simple manipulation with substituting (13) and (15) into μ. ■

Corollary 1 states that we have a degree-of-freedom (DoF) loss for $\mathcal{R}_{sum}(\mathcal{F}_k)$ compared to $\mathcal{R}^\star_{sum}$ for the training beam design B.1. On the other hand, when the training beam B.2 is used, we have full DoF but an array-gain loss of ξ^2. This implies that when we use narrow training beams, equally-spaced entire-angle covering beams have better performance than the consecutive beams covering a small range of angles if K users are spread over the entire angle domain. The performance difference between B.1 and B.2 results from the fact that the equally-spaced beams additionally exploit the major side lobes to estimate to some extent the channel components between the boresights of two adjacent beams, whereas the consecutive beams do not. In the case of B.3, we have full DoF and the array-gain loss of only ξ, which is smaller than ξ^2 since $0 < \xi < 1$.

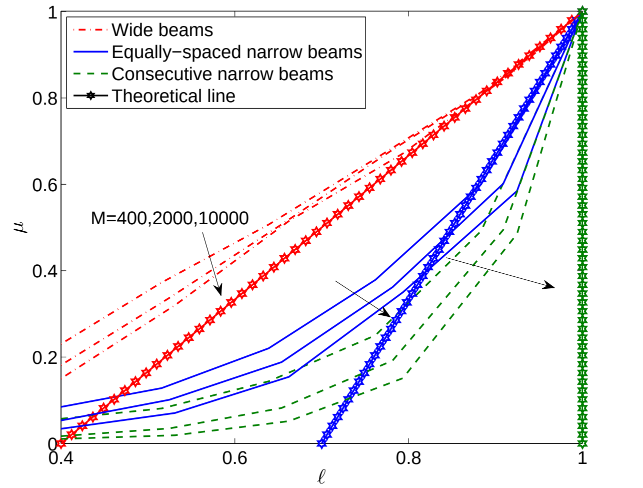

Fig. 2. μ versus ℓ when $q = 0.4$.

Theorem 2 gives a rough answer to our intial question "what is the performance loss of partial training relative to full training in mmWave MIMO downlink?" since the ratio μ of $\mathcal{R}_{sum}(\mathcal{F}_k)$ to $\mathcal{R}^\star_{sum}$ is quantified as an explicit function of q, ℓ and the type of training beams. For B.3, the ratio μ linearly increases from 0 to 1 as ℓ increases from q to 1. For B.2, μ starts to increase from 0 at $\ell = \frac{1+q}{2}$, which is larger than q for $q < 1$. In the case of B.1, the asymptotic ratio μ is zero for $\ell \in (q, 1)$.

IV. Numerical Results

In this section, we provide a numerical result to validate our asymptotic results in the previous section. All the expectations were performed over 100 channel realizations and we set $P_t = 1$, $N = 4$ and $M_{RF} = K$. Fig. 2 shows the ratio $\mu(= \frac{\mathcal{R}_{sum}(\mathcal{F}_k)}{\mathcal{R}^\star_{sum}})$ for each type of training beams versus ℓ for $M = 400, 2000, 10000$ when $q = 0.4$. Since we modeled $K = \lfloor M^q \rfloor$ in this simulation, we have $K = 10, 20$, and 39 for $M = 400, 2000$, and 10000, respectively. We computed $\mathcal{R}_{sum}(\mathcal{F}_k)$ and $\mathcal{R}^\star_{sum}$ by simulating the two-stage feedback-based hybrid precoding design method proposed in [7] with the assumption of perfect CSI at the second stage. It is seen that the curve of μ of each type of training beams gradually converges to each theoretical line as M increases. Note that there exists some gap between each theoretical asymptotic line and the finite-sample result. This results from the slow rate of convergence.

V. Conclusions

In this paper, we have considered mmWave MU-MIMO downlink systems with hybrid A/D beamforming and examined the impact of training on the rate performance of systems. We have quantified the performance loss w.r.t. the number S of training beams in large-scale mmWave MU-MIMO for three meaningful different types of training beams under the UR-SP channel model that captures mmWave MIMO channels. The provided results in this paper give some insights into the impact of training design in mmWave MU-MIMO and more efficient training beam design is left as future work.

REFERENCES

[1] Z. Pi and F. Khan, "An introduction to millimeter-wave mobile broadband systems," *IEEE Commun. Mag.*, vol. 49, pp. 101–107, Jun. 2011.

[2] S. Hur, T. Kim, D. Love, J. Krogmeier, T. Thomas, A. Ghosh, "Millimeter wave beamforming for wireless backhaul and access in small cell networks," *IEEE Trans. Commun.*, vol. 61, pp. 4391–4403, Oct. 2013.

[3] O. E. Ayach, S. Rajagopal, S. Abu-Surra, P. Zhouyue, and R. W. Heath, "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Trans. Wirelee Commun.*, vol. 13, pp. 1499 – 1513, Mar. 2014.

[4] A. Alkhateeb, O. E. Ayach, G. Leus, and R. Heath , "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE J. Sel. Topics Signal Process.*, vol. 8, pp. 831 – 846, Oct. 2014.

[5] J. Seo, Y. Sung, G. Lee, and D. Kim, "Training beam sequence design for millimeter-wave MIMO systems: A POMDP framework," *IEEE Trans. Signal Process.*, vol. 64, pp. 1228 – 1242, Mar. 2016.

[6] J. Lee, G. Gil, and Y. Lee, "Channel estimation via orthogonal matching pursuit for hybrid MIMO systems in millimeter wave communications," *to appear in IEEE Trans. Commun.*, 2016.

[7] A. Alkhateeb, G. Leus, and R. Heath , "Limited feedback hybrid precoding for multi-user millimeter wave systems," *IEEE Trans. Wireless Commun.*, vol. 14, pp. 6481–6494, Nov. 2015.

[8] J. Choi, "Beam selection in mm-wave multiuser MIMO systems using compressive sensing," *IEEE Trans. Commun.*, vol. 63, pp. 2936–2947, Aug. 2016.

[9] G. Lee, Y. Sung, and J. Seo, "Randomly-directional beamforming in millimeter-wave multi-user MISO downlink," *IEEE Trans. Wireless Commun.*, vol. 15, pp. 1086–1100, Feb. 2016.

[10] G. Lee, Y. Sung, and M. Kountouris, "On the performance of random beamforming in sparse millimeter wave channels," *IEEE J. Sel. Topics Signal Process.*, vol. 10, pp. 560–575, Apr. 2016.

[11] A. M. Sayeed, "Deconstructing multiantenna fading channels," *IEEE Trans. Signal Process.*, vol. 50, pp. 2563 – 2579, Oct. 2002.

[12] A. M. Sayeed and V. Raghavan, "Maximizing MIMO capacity in sparse multipath with reconfigurable antenna arrays," *IEEE J. Sel. Topics Signal Process.*, vol. 1, pp. 156 – 166, Jun. 2007.

[13] Y. Tsang, A. Poon, and S. Addepalli, "Coding the beams: Improving beamforming training in mmWave communication system," in *Proc. IEEE Globecom*, Dec. 2011.

[14] H. L. Van Trees, *Optimum Array Processing*. New York: John Wiley & Sons, 2002.

[15] A. Sayeed and J. Brady, "Beamspace MIMO for high-dimensional multiuser communication at millimeter-wave frequencies," in *Proc. IEEE Global Telecommun. Conf. (Globecom)*, pp. 3679 – 3684, Dec. 2013.

[16] H. Q. Ngo, E. G. Larsson, and T. L Marzetta, "Aspects of favorable propagation in massive MIMO," in *Proc. IEEE EUSIPCO 2014*, pp. 76 – 80, Sep. 2014.