Beam Training with Limited Feedback for Multiuser mmWave Massive MIMO

Zhaohui Li* and Chenhao Qi†
*School of Cyber Science and Engineering, Southeast University, Nanjing, China
†School of Information Science and Engineering, Southeast University, Nanjing, China
Email: qch@seu.edu.cn

GLOBECOM 2020 - 2020 IEEE Global Communications Conference | 978-1-7281-8298-8/20/$31.00 ©2020 IEEE | DOI: 10.1109/GLOBECOM42002.2020.9348205

***Abstract*—Different from the existing hierarchical beam training schemes which typically require a feedback in each layer of the codebook to indicate the best codeword for the base station (BS), beam training in this work only needs two feedbacks in total. The proposed beam training scheme includes two stages and requires only one feedback in each stage by using the designed hierarchical codebook. In the first stage, beam training for the top layer of our designed codebook is performed to narrow the search range of the channel angle of departure (AoD). In the second stage, based on the feedback information from the first stage, beam training for the other layers of our designed codebook is performed, where the BS can obtain an estimate of the channel AoD. Simulation results show that the performance of the proposed beam training scheme with limited feedback can approach that of time-division multiple access (TDMA) hierarchical beam training scheme with much fewer feedbacks.**

***Index Terms*—Millimeter wave (mmWave) communications, massive MIMO, beam training, hierarchical codebook**

I. Introduction

Millimeter wave (mmWave) communication is a promising technology to tackle the increasingly scarce spectrum resources owing to its wide bandwidth from 30GHz to 300GHz [1]. However, mmWave communication has a shortcoming of severe path loss. To deal with it, directional beamforming with massive MIMO antenna arrays has been introduced, where the antenna arrays can be encapsulated in a small size of circuits due to the short wavelength at mmWave frequency. Note that the fully digital beamforming used in the conventional MIMO working at sub 6GHz is impractical for mmWave massive MIMO, as it needs a large number of expensive radio frequency (RF) chains. To approach the sum-rate performance of fully digital beamforming with much lower hardware cost, hybrid beamforming with only a few RF chains connected to the antenna arrays via phase shifters is proposed [2].

To obtain the channel state information (CSI) required by the hybrid beamforming, conventional entry-wise estimation of mmWave massive MIMO channel matrix requires a huge temporal, frequency or spatial resources [3], [4]. Since the mmWave channel matrix can be modeled as a weighted summation of the product of a transmit channel steering vector and a receive channel steering vector, a better approach to obtain the CSI is using beam training based on a predefined codebook for both the transmitter and receiver [5], [6], where each codeword of the codebook has a similar representation as the channel steering vector. Then the CSI acquisition is essentially to find a pair of transmit and receive codewords best matched with the mmWave massive MIMO channel. Compared to beam sweeping that exhaustively tests all pairs of transmit and receive codewords, the hierarchical beam training scheme based on hierarchical codebooks can significantly reduce the beam training overhead. A hierarchical codebook typically includes a number of layers of codebooks with multi-resolution codewords, where the beam coverage of a codeword at a high layer covers that of several codewords at lower layers [7], [8]. For single-user mmWave massive MIMO, a hierarchical codebook design method named beam widening via single-RF subarray (BMW-SS) is proposed, where the beam width of different codewords can be flexibly adjusted by dividing the large antenna array into several small sub-arrays and switching on-off some antennas [4]. In fact, the hierarchical beam training based on BMW-SS can be extended from the single-user scenario to the multiuser scenario, where the base station (BS) performs the beam training for each user in turn with time-division multiple access (TDMA). But the beam training overhead may grow linearly with the number of users. To reduce the overhead, a simultaneous multiuser beam training scheme is proposed, where each layer excluding the bottom layer has only two multi-mainlobe codewords no matter how many users the BS serves [9]. Note that the hierarchical beam training schemes typically require a feedback from a user equipment (UE) to the BS to indicate the best codeword, once finishing the beam training for each layer of the hierarchical codebook. Since the feedback requires non-trivial uplink and downlink coordination between the UEs and the BS [10], the beam training schemes with limited feedback are worthy of further studies.

In this paper, a beam training scheme with limited feedback is proposed. The scheme includes two stages and requires only one feedback in each stage by using the designed hierarchical codebook. In the first stage, beam training for the top layer of our designed codebook is performed to narrow the search range of the channel angle of departure (AoD). In the second stage, based on the feedback information from the first stage, beam training for the other layers of our designed codebook is performed, where the BS can obtain an estimate of the channel AoD. The scheme only needs totally two feedbacks for each user, no matter how many layers the

978-1-7281-8298-8/20/$31.00 ©2020 IEEE

hierarchical codebook has.

Notations: Symbols for matrices (upper case) and vectors (low case) are in boldface. The symbols $\mathbb{E}\{\cdot\}$, $\boldsymbol{I}$, $\mathcal{CN}$ and $\mathbb{C}$ denote the statistical expectation, identity matrix, complex Gaussian distribution and complex number set, respectively. $\angle(a)$ and $|a|$ denote the angle and amplitude of a complex number a. $[\boldsymbol{v}]_m$ denotes the mth entry of a vector $\boldsymbol{v}$. For a matrix $\boldsymbol{M}$, the symbols $[\boldsymbol{M}]_{:,n}$, $\boldsymbol{M}^{\mathrm{T}}$, $\boldsymbol{M}^{\mathrm{H}}$, $\|\boldsymbol{M}\|_2$ and $\|\boldsymbol{M}\|_{\mathrm{F}}$ denote its nth column, transpose, conjugate transpose, ℓ_2-norm and Frobenius norm, respectively. The symbol $\odot$ denotes the entry-wise product between two matrices.

II. System Model

Consider a multiuser mmWave massive MIMO system with a BS and K UEs. The BS employs the hybrid beamforming structure to connect its N_{RF} RF chains to N_{BS} antennas ($N_{\mathrm{BS}} \gg N_{\mathrm{RF}} \geq K$), while each UE connects its single RF chain to N_{UE} ($N_{\mathrm{BS}} \geq N_{\mathrm{UE}}$) antennas using analog combining. Note that both N_{BS} and N_{UE} are in an integer power of two. The antennas at both the BS and UEs are placed in uniform linear arrays (ULAs) with neighbouring two antennas spaced in half wavelength.

The received signal by the kth UE for $k = 1, 2, \ldots, K$ is expressed as

$$y_k = \boldsymbol{w}_k^{\mathrm{H}} \boldsymbol{H}_k \boldsymbol{F}_{\mathrm{RF}} \boldsymbol{F}_{\mathrm{BB}} \boldsymbol{x} + \boldsymbol{w}_k^{\mathrm{H}} \boldsymbol{n}_k, \tag{1}$$

where $\boldsymbol{F}_{\mathrm{RF}} \in \mathbb{C}^{N_{\mathrm{BS}} \times N_{\mathrm{RF}}}$ and $\boldsymbol{F}_{\mathrm{BB}} \in \mathbb{C}^{N_{\mathrm{RF}} \times K}$ denote the analog beamformer and the digital beamformer of the BS, respectively, and satisfy $\|\boldsymbol{F}_{\mathrm{RF}} \boldsymbol{F}_{\mathrm{BB}}\|_F^2 = K$; $\boldsymbol{w}_k \in \mathbb{C}^{N_{\mathrm{UE}}}$ denotes the analog combiner of the kth UE; $\boldsymbol{H}_k \in \mathbb{C}^{N_{\mathrm{UE}} \times N_{\mathrm{BS}}}$ denotes the downlink mmWave MIMO channel matrix; $\boldsymbol{x} \in \mathbb{C}^K$ denotes the transmit signal such that $\mathbb{E}\{\boldsymbol{x}\boldsymbol{x}^{\mathrm{H}}\} = \frac{P}{K} \boldsymbol{I}_K$ with P representing the transmit power; and $\boldsymbol{n}_k \in \mathbb{C}^{N_{\mathrm{UE}}}$ denotes the Gaussian noise vector obeying $\boldsymbol{n}_k \sim \mathcal{CN}(0, \sigma^2 \boldsymbol{I}_{N_{\mathrm{UE}}})$.

According to the widely used Saleh Valenzuela model [1], $\boldsymbol{H}_k$ can be expressed as

$$\boldsymbol{H}_k = \sqrt{\frac{N_{\mathrm{BS}} N_{\mathrm{UE}}}{L}} \sum_{l=1}^{L} \lambda_l \boldsymbol{a}(N_{\mathrm{UE}}, \theta_{\mathrm{UE},k}^l) \boldsymbol{a}^{\mathrm{H}}(N_{\mathrm{BS}}, \theta_{\mathrm{BS},k}^l), \tag{2}$$

where L denotes the number of multipath. λ_l, $\theta_{\mathrm{UE},k}^l$ and $\theta_{\mathrm{BS},k}^l$ denote the complex channel coefficient, the channel angle of arrival (AoA), and the channel AoD of the lth path for $l = 1, 2, \ldots, L$, respectively. Additionally, we have $\theta_{\mathrm{UE},k}^l \in [-1, 1]$ and $\theta_{\mathrm{BS},k}^l \in [-1, 1]$. The channel steering vector is defined as

$$\boldsymbol{a}(N, \phi) \triangleq \frac{1}{\sqrt{N}} \left[1, e^{j\pi\phi}, \ldots, e^{j(N-1)\pi\phi}\right]^{\mathrm{T}}, \tag{3}$$

where N denotes the antenna number, and ϕ denotes the channel AoA or AoD.

Suppose all the UEs employ the same codebook. Then the codebooks at the BS and the UEs are denoted as $\mathcal{F} = \{\boldsymbol{f}_c^1, \boldsymbol{f}_c^2, \ldots, \boldsymbol{f}_c^{N_{\mathrm{BS}}}\}$ and $\mathcal{W} = \{\boldsymbol{w}_c^1, \boldsymbol{w}_c^2, \ldots, \boldsymbol{w}_c^{N_{\mathrm{UE}}}\}$, respectively, where

$$\boldsymbol{f}_c^n = \boldsymbol{a}(N_{\mathrm{BS}}, -1 + (2n-1)/N_{\mathrm{BS}}),\ n = 1, 2, \ldots, N_{\mathrm{BS}}, \tag{4a}$$

$$\boldsymbol{w}_c^m = \boldsymbol{a}(N_{\mathrm{UE}}, -1 + (2m-1)/N_{\mathrm{UE}}),\ m = 1, 2, \ldots, N_{\mathrm{UE}}. \tag{4b}$$

During the beam training for the kth UE for $k = 1, 2, \ldots, K$, we aim to find a pair of $\boldsymbol{f}_k \triangleq [\boldsymbol{F}_{\mathrm{RF}}]_{:,k} \in \mathcal{F}$ and $\boldsymbol{w}_k \in \mathcal{W}$ best matched with $\boldsymbol{H}_k$, which can be formulated as the following optimization problem [11]

$$\max_{\boldsymbol{f}_k, \boldsymbol{w}_k} \left|\boldsymbol{w}_k^{\mathrm{H}} \boldsymbol{H}_k \boldsymbol{f}_k\right|,\ k = 1, 2, \ldots, K, \tag{5a}$$

$$\text{s.t.}\quad \boldsymbol{f}_k \in \mathcal{F}, \boldsymbol{w}_k \in \mathcal{W}. \tag{5b}$$

The straightforward method to solve (5) is beam sweeping [12], which exhaustively tests all pairs of $\boldsymbol{f}_k$ and $\boldsymbol{w}_k$ to find the best one. Assume the test of each pair of codewords takes one time slot. The training overhead for beam sweeping is $N_{\mathrm{BS}} N_{\mathrm{UE}}$ time slots. Since each UE needs to report the best codeword for the BS, beam sweeping needs K feedbacks in total after finishing beam training. Although the beam sweeping only needs a small number of feedbacks, it has a large training overhead.

To reduce the beam training overhead, hierarchical beam training (HBT) [4] is adopted based on the idea of binary search to divide the search range of the channel AoD layer by layer. Specifically, it starts with low-resolution codewords to search in a wide range of the channel AoD and then continuously improves codeword resolution to refine the search based on the feedback information. In the multiuser scenario, we can perform the HBT for each user in turn in TDMA manner, which is called TDMA-HBT. Then TDMA-HBT scheme for K users needs the beam training overhead of $2K(\log_2 N_{\mathrm{BS}} + \log_2 N_{\mathrm{UE}})$ in total. Since each UE needs to report the best codeword to the BS after finishing the beam training of each layer of the hierarchical codebook, the TDMA-HBT scheme needs $K \log_2 N_{\mathrm{BS}}$ feedbacks in total. In order to further reduce the training overhead, a simultaneous multiuser beam training (SMBT) scheme is proposed [9], where each layer excluding the bottom layer has only two multi-mainlobe codewords no matter how many users the BS serves. The SMBT scheme needs beam training overhead of $2(K + \log_2 N_{\mathrm{BS}} N_{\mathrm{UE}} - 1)$ and requires $K(\log_2 N_{\mathrm{BS}} - 1)$ feedbacks. Note that either TDMA-HBT or SMBT is based on hierarchical codebooks, which require feedback in each layer of the hierarchical codebook. Since the feedback requires non-trivial uplink and downlink coordination between the UEs and the BS [10], we will consider the beam training with limited feedback.

III. Hierarchical Codebook-based Beam Training with Limited Feedback

In this section, we first design a hierarchical codebook by proposing a codeword design algorithm. Based on the designed hierarchical codebook, we then propose a beam

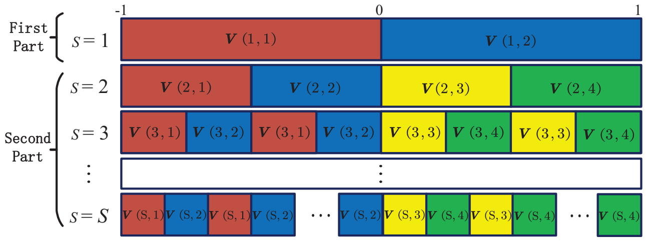

Fig. 1. Illustration of designed hierarchical codebook $\boldsymbol{V}$ with $N_1 = 2$.

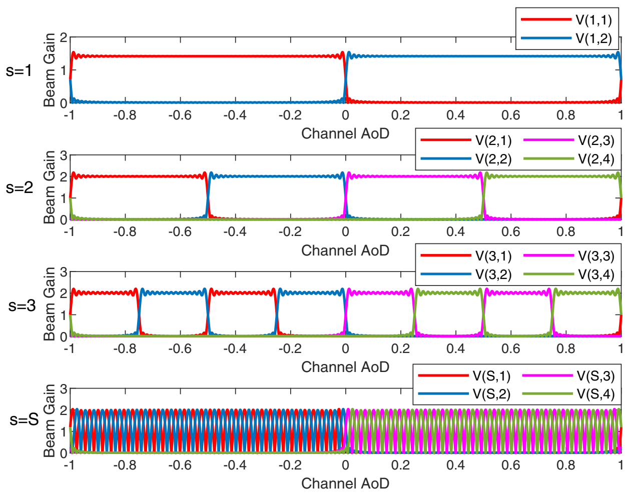

Fig. 2. Beam pattern of designed hierarchical codebook $\boldsymbol{V}$ with $N_1 = 2$.

training scheme, which can significantly reduce the number of feedbacks compared with the existing TDMA-HBT scheme. Note that the designed hierarchical codebook is used at BS, while the codebooks at UEs are designed using the existing method [4].

A. *Hierarchical Codebook Design*

As shown in Fig. 1, we design a hierarchical codebook $\boldsymbol{V}$ including S layers of codebooks. We divide $\boldsymbol{V}$ into two parts, where the first part is essentially the top layer of $\boldsymbol{V}$ and the second part includes the other layers of $\boldsymbol{V}$. The number of codewords in the top layer is denoted as N_1, which is determined by the beam gain related to the signal coverage of the BS, e.g., we may enlarge N_1 and use N_1 narrow beams to improve the signal coverage. Note that N_1 is normally set to be an integer power of two. We give an example of $\boldsymbol{V}$ with $N_1 = 2$, which is illustrated in Fig. 1. Since the second part of $\boldsymbol{V}$ includes $\log_2(N_{\rm BS}/N_1)$ layers of codebooks, we have

$$S = \log_2(N_{\rm BS}/N_1) + 1. \tag{6}$$

The number of codewords in each layer of the second part is $2N_1$. In particular, we design the codewords in the layers from $s = 3$ to $s = S$ to be multi-mainlobe codewords [13], where s denotes the index of the layers of $\boldsymbol{V}$. In [13], a hierarchical codebook is designed in an *adaptive* manner, where the codewords in the current layer are designed according to the feedbacks from the UEs on the beam training results of the previous layer. Different from [13], in this work we design $\boldsymbol{V}$ to be a *fixed* hierarchical codebook.

Algorithm 1 Codeword Design Algorithm

Input: $\boldsymbol{A}$ and $\widehat{\boldsymbol{g}}$.
Initialization: Set $\boldsymbol{t} = \underbrace{[1, 1, \ldots, 1]}_{U}$.
while *Stop Condition* is not satisfied **do**
  Given $\boldsymbol{t}$, obtain $\widehat{\boldsymbol{v}}$ via (15).
  Give $\widehat{\boldsymbol{v}}$, update $\boldsymbol{t}$ via (16).
end while
Output: $\widehat{\boldsymbol{v}}$.

In the following, we design the first codeword of each layer of $\boldsymbol{V}$, i.e., $\boldsymbol{V}(s, 1)$ for $s = 1, 2, \ldots, S$. Based on $\boldsymbol{V}(s, 1)$, we can fast obtain the other codewords in the sth layer by shifting the center of the beam coverage of $\boldsymbol{V}(s, 1)$. To ease the notation, we redefine $\boldsymbol{v} \triangleq \boldsymbol{V}(s, 1) \in \mathbb{C}^{N_{\rm BS}}$. The beam gain of $\boldsymbol{v}$ along θ is defined as a function of $\boldsymbol{v}$ and θ as

$$g(\boldsymbol{v}, \theta) \triangleq \sqrt{N_{\rm BS}}\boldsymbol{a}(N_{\rm BS}, \theta)^{\rm H}\boldsymbol{v} = \sum_{i=1}^{N_{\rm BS}} [\boldsymbol{v}]_i e^{-j\pi(i-1)\theta} \tag{7}$$

for $\theta \in [-1, 1]$. The beam coverage of $\boldsymbol{v}$ is denoted as $\mathcal{B}_{\boldsymbol{v}}$. By temporarily omitting the power constraint $\|\boldsymbol{v}\|_2 = 1$ and assuming the normalized absolute beam gain [14], we can obtain

$$|g(\boldsymbol{v}, \theta)| = \begin{cases} 1, & \theta \in \mathcal{B}_{\boldsymbol{v}}, \\ 0, & \theta \notin \mathcal{B}_{\boldsymbol{v}}. \end{cases} \tag{8}$$

We define

$$\boldsymbol{A} \triangleq \sqrt{N_{\rm BS}}\big[\boldsymbol{a}(N_{\rm BS}, \theta_1), \boldsymbol{a}(N_{\rm BS}, \theta_2), \ldots, \boldsymbol{a}(N_{\rm BS}, \theta_U)\big], \tag{9}$$

as a matrix made up of $U (U \geq N_{\rm BS})$ channel steering vectors, where

$$\theta_u = -1 + (2u-1)/U, \ u = 1, 2, \ldots, U \tag{10}$$

is the quantized channel AoD. In fact, we use U channel steering vectors to equally sample the angular space $[-1, 1]$.

Then the codeword design problem can be modeled as

$$\min_{\boldsymbol{v}} \| \boldsymbol{A}^{\rm H}\boldsymbol{v} - \boldsymbol{g} \|_2^2, \tag{11}$$

where $\boldsymbol{g} \in \mathbb{C}^U$ denotes an objective beam gain vector with

$$[\boldsymbol{g}]_u = g(\boldsymbol{v}, \theta_u), \ u = 1, 2, \ldots, U. \tag{12}$$

As U grows to be infinity, $\boldsymbol{g}$ will approach the ideal beam gain in continuous angular space.

To solve (11), we propose a codeword design algorithm based on the cyclic algorithm [15].

We define a normalized absolute beam gain vector as $\widehat{\boldsymbol{g}} \in \mathbb{C}^U$, where

$$[\widehat{\boldsymbol{g}}]_u = |g(\boldsymbol{v}, \theta_u)|, \ u = 1, 2, \ldots, U. \tag{13}$$

Note that different phases of $g(\boldsymbol{v}, \theta)$ do not change $|g(\boldsymbol{v}, \theta)|$ in (8). Therefore, we introduce a phase vector $\boldsymbol{t} \in \mathbb{C}^U$, where $\big|[t]_u\big| = 1$ for $u = 1, 2, \ldots, U$. We initialize $\boldsymbol{t}$ to be a vector with all the entries being one and substitute $\boldsymbol{g}$ in (11) by $\widehat{\boldsymbol{g}} \odot \boldsymbol{t}$, obtaining

$$\min_{\boldsymbol{v}} \parallel \boldsymbol{A}^{\mathrm{H}} \boldsymbol{v} - \widehat{\boldsymbol{g}} \odot \boldsymbol{t} \parallel_2^2, \tag{14}$$

where the least squares (LS) solution of (14) can be expressed as

$$\widehat{\boldsymbol{v}} = (\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}})^{-1} \boldsymbol{A} (\widehat{\boldsymbol{g}} \odot \boldsymbol{t}). \tag{15}$$

Then we can update $\boldsymbol{t}$ by

$$[\boldsymbol{t}]_u = e^{j\angle([\boldsymbol{A}^{\mathrm{H}}\widehat{\boldsymbol{v}}]_u)},\ u = 1, 2, \ldots, U. \tag{16}$$

We iteratively perform (15) and (16) until a *Stop Condition* is satisfied. The *Stop Condition* can be set that the a prespecified maximum number of iterations is reached or the ℓ_2 norm of $\widehat{\boldsymbol{v}}$ in two consecutive iterations is smaller than a prespecified threshold.

The detailed steps of the algorithm is summarized in **Algorithm** 1. It is seen that once $\mathcal{B}_{\boldsymbol{v}}$ in (8) and U are given, we can run **Algorithm** 1 to design $\boldsymbol{v}$. When further considering hardware constraints including the number of RF chains and the resolution of phase shifters, we may resort to the existing practical codeword design schemes [14].

Now we give an example for $N_1 = 2$. For the top layer where $s = 1$, we set $\mathcal{B}_{\boldsymbol{V}(1,1)} = [-1, 0]$ to design $\boldsymbol{V}(1,1)$. Based on $\boldsymbol{V}(1,1)$, we can fast obtain $\boldsymbol{V}(1,2)$ by shifting the center of the beam coverage of $\boldsymbol{V}(1,1)$. For the other layers where $s = 2, 3, \ldots, S$, each layers have four codewords, where each codeword has 2^{s-2} mainlobes and the beam width of each mainlobe is $1/2^{s-1}$. In fact, each codeword in the second layer has only one mainlobe, while each codeword in the layers with $s \geq 3$ has multiple mainlobes and the width between the center of two neighbouring mainlobes is $1/2^{s-1}$. In this way, we set $\mathcal{B}_{\boldsymbol{V}(2,1)} = [-1, -0.5]$ to design $\boldsymbol{V}(2,1)$. We set $\mathcal{B}_{\boldsymbol{V}(3,1)} = [-1, -0.75] \cup [-0.5, -0.25]$ to design $\boldsymbol{V}(3,1)$. It is seen that $\boldsymbol{V}(3,1)$ has two mainlobes, where the beam coverage of one mainlobe is $[-1, -0.75]$ and that of the other mainlobe is $[-0.5, -0.25]$. Based on $\boldsymbol{V}(3,1)$, we can fast obtain $\boldsymbol{V}(3,2)$, $\boldsymbol{V}(3,3)$ and $\boldsymbol{V}(3,4)$ by shifting the center of the beam coverage of $\boldsymbol{V}(3,1)$. The codewords in the other layers can be similarly designed. Note that **Algorithm** 1 can be used to design codewords with single mainlobe, as well as those with multiple mainlobes.

In Fig. 2, we give the beam pattern of the codewords designed by **Algorithm** 1, where $N_1 = 2$ and $U = 16384$. It is seen that the beam coverage of $\boldsymbol{V}(1,1)$ can cover that of $\boldsymbol{V}(s,1)$ and $\boldsymbol{V}(s,2)$ for $s = 2, 3, \ldots, S$. Similarly, the beam coverage of $\boldsymbol{V}(1,2)$ can cover that of $\boldsymbol{V}(s,3)$ and $\boldsymbol{V}(s,4)$ for $s = 2, 3, \ldots, S$. It is also observed that the beam width of any codeword in the second part is the same, where the beam width of a multi-mainlobe codeword is defined as the summation of beam width of all the mainlobes.

Algorithm 2 Beam Training with Limited Feedback

```
Input: V.
(Stage 1)
Perform beam training using V(1,1), V(1,2), ...,
  V(1,N_1), respectively.
The UE feedbacks I_1 to the BS.
(Stage 2)
for s = 2,3,...,S do
    Perform beam training using V(s,2I_1 − 1) and
    V(s,2I_1), respectively.
    The UE records "0" or "1" in a sequence.
end for
The UE feedbacks the binary sequence to the BS.
Calculate I via (18).
Output: I.
```

B. Beam Training with Limited Feedback

Based on the designed hierarchical codebook in the previous subsection, we propose a beam training scheme with limited feedback. The scheme includes two stages and requires only one feedback in each stage. In the first stage, beam training for the top layer of our designed codebook is performed to narrow the search range of the channel AoD. In the second stage, based on the feedback information from the first stage, beam training for the other layers of our designed codebook is performed, where the BS can obtain an estimate of the channel AoD.

The BS sequentially performs the beam training with all the K UEs. In the following, we present the beam training of the BS with the kth UE for $k = 1, 2, \ldots, K$.

In the first stage, we perform beam training using the codewords in the top layer of $\boldsymbol{V}$, where all N_1 codewords at the BS are tested. Then the kth UE feedbacks to the BS the index of the best BS codeword, which is denoted as an integer I_1 with $I_1 \in \{1, 2, \ldots, N_1\}$ occupying $\lceil \log_2(N_1) \rceil$ bits. The BS can narrow the search of the channel AoD.

In the second stage, we perform beam training using the second part of $\boldsymbol{V}$. For the sth layer, $s = 2, 3, \ldots, S$, the BS sequentially transmits the beam generated by $\boldsymbol{V}(s, 2I_1 - 1)$ and $\boldsymbol{V}(s, 2I_1)$. If the kth UE finds that $\boldsymbol{V}(s, 2I_1 - 1)$ results in higher received signal power than $\boldsymbol{V}(s, 2I_1)$, the kth UE records a new bit "0"; otherwise, the kth UE records a new bit "1". The new bit is always placed at the end of the old bits that are recorded during the beam training at the previous layers of $\boldsymbol{V}$. Note that for each layer in the second part of $\boldsymbol{V}$, the BS only tests two codewords. In this way, the kth UE records a binary sequence with $S - 1$ bits, which are then fed back to the BS and the BS can obtain an estimate of the channel AoD. Suppose the value of this binary sequence equals I_2 in decimal. We can obtain a solution to (5) as

$$\widehat{\boldsymbol{f}}_k = \boldsymbol{f}_c^I \tag{17}$$

where

$$I = N_{\mathrm{BS}}(I_1 - 1)/N_1 + I_2 + 1. \tag{18}$$

TABLE I
OVERHEAD COMPARISONS

Schemes	Training overhead	The number of feedbacks
Our scheme	$2K\big(N_1 + \log_2(N_{\rm UE}) + \log_2(N_{\rm BS}/N_1) - 1\big)$	$2K$
Beam sweeping	$N_{\rm BS}N_{\rm UE}$	K
TDMA-HBT	$2K(\log_2 N_{\rm BS} + \log_2 N_{\rm UE})$	$K\log_2 N_{\rm BS}$
SMBT	$2(K + \log_2 N_{\rm BS}N_{\rm UE} - 1)$	$K(\log_2 N_{\rm BS} - 1)$

We denote the designed analog beamformer as $\widehat{\boldsymbol{F}}_{\rm RF}$, where

$$\big[\widehat{\boldsymbol{F}}_{\rm RF}\big]_{:,k} = \widehat{\boldsymbol{f}}_k,\ k = 1, 2, \ldots, K. \tag{19}$$

The procedures of the proposed beam training scheme are summarized in **Algorithm** 2. Note that the existing hierarchical beam training requires a feedback at each layer of the hierarchical codebook [4], which results in totally $K\log_2 N_{\rm BS}$ feedbacks. Our scheme only requires totally $2K$ feedbacks, which is very limited compared to the existing scheme.

C. Digital beamforming

After the beam training, the obtained analog combiner for the kth UE is denoted as $\widehat{\boldsymbol{w}}_k \in \mathcal{W},\ k = 1, 2, \ldots, K$. To estimate the effective channel $\boldsymbol{H}_e$ that can be denoted as

$$\boldsymbol{H}_e \triangleq \begin{bmatrix} \widehat{\boldsymbol{w}}_1^{\rm H}\boldsymbol{H}_1\widehat{\boldsymbol{f}}_1 & \widehat{\boldsymbol{w}}_1^{\rm H}\boldsymbol{H}_1\widehat{\boldsymbol{f}}_2 & \cdots & \widehat{\boldsymbol{w}}_1^{\rm H}\boldsymbol{H}_1\widehat{\boldsymbol{f}}_K \\ \widehat{\boldsymbol{w}}_2^{\rm H}\boldsymbol{H}_2\widehat{\boldsymbol{f}}_1 & \widehat{\boldsymbol{w}}_2^{\rm H}\boldsymbol{H}_2\widehat{\boldsymbol{f}}_2 & \cdots & \widehat{\boldsymbol{w}}_2^{\rm H}\boldsymbol{H}_2\widehat{\boldsymbol{f}}_K \\ \vdots & \cdots & \ddots & \vdots \\ \widehat{\boldsymbol{w}}_K^{\rm H}\boldsymbol{H}_2\widehat{\boldsymbol{f}}_1 & \widehat{\boldsymbol{w}}_K^{\rm H}\boldsymbol{H}_K\widehat{\boldsymbol{f}}_2 & \cdots & \widehat{\boldsymbol{w}}_K^{\rm H}\boldsymbol{H}_2\widehat{\boldsymbol{f}}_K \end{bmatrix}, \tag{20}$$

we need to perform baseband channel estimation. Note that the dimension of $\boldsymbol{H}_e$ is much smaller than $\boldsymbol{H}_k$. Compared to directly estimating RF frontend channel matrix $\boldsymbol{H}_k$, the baseband channel estimation is much easier. Finally, the digital beamformer under zero forcing (ZF) criterion can be denoted as

$$\widehat{\boldsymbol{F}}_{\rm BB} = \boldsymbol{H}_e^{\rm H}(\boldsymbol{H}_e\boldsymbol{H}_e^{\rm H})^{-1}. \tag{21}$$

D. Overhead Analysis

Since the number of feedbacks is an important performance metric, we compare it in Table I for different schemes including the beam sweeping, TDMA-HBT, SMBT scheme, and our scheme. In fact, we have already provided these data in the last paragraph of **Section** II. For example, if $N_{\rm BS} = 128$, $N_{\rm UE} = 16$, $N_1 = 2$ and $K = 4$, the number of feedbacks of beam sweeping, our scheme, SMBT and TDMA-HBT are 4, 8, 24 and 28, respectively. Compared to SMBT and TDMA-HBT, our scheme can achieve 67% and 71% reduction in the number of feedbacks, respectively.

We also compare the beaming training overhead for these schemes in Table I. For our scheme, we use a hierarchical codebook with $\log_2(N_{\rm UE})$ layers at each UE and we start the beam training from the first layer of the UE, which is the same as TDMA-HBT and SMBT. At the top layer of $\boldsymbol{V}$, all the $2N_1$ transmit-and-receive beam pairs between the N_1 codewords at the BS and two codewords at the UE are sequentially tested. Suppose $S > \log_2(N_{\rm UE})$ with S defined in (6). During the beam training from $s = 2$ to $s = \log_2(N_{\rm UE})$, we need to test 4 beam pairs for each layer, which results in totally $4(\log_2(N_{\rm UE}) - 1)$ tests. After that, the best codeword at the UE is determined and we only need to test the codewords at the BS. For the remaining layers from $s = \log_2(N_{\rm UE}) + 1$ to $s = S$, we need to test 2 codewords for each layer of $\boldsymbol{V}$, which results in totally $2(S - \log_2(N_{\rm UE}))$ tests. Considering totally K UEs, the beam training overhead for our scheme is

$$\begin{aligned} &K\left(2N_1 + 4(\log_2(N_{\rm UE}) - 1) + 2\big(S - \log_2(N_{\rm UE})\big)\right) \\ =&2K\big(N_1 + \log_2(N_{\rm UE}) + \log_2(N_{\rm BS}/N_1) - 1\big). \end{aligned} \tag{22}$$

The training overhead of the other schemes is provided in [13]. For example, if $N_{\rm BS} = 128$, $N_{\rm UE} = 16$, $N_1 = 2$ and $K = 4$, the training overhead of beam sweeping, our scheme, SMBT and TDMA-HBT are 2048, 88, 28 and 88, respectively. Since the SMBT performs simultaneous beam training with all the UEs, it achieves the smallest training overhead. Our scheme has the same training overhead as TDMA-HBT.

IV. SIMULATION RESULTS

Now we evaluate the performance of the propose beam training scheme. The BS equipped with $N_{\rm BS} = 128$ antennas serves $K = 4$ UEs, where each UE is equipped with $N_{\rm UE} = 16$ antennas. We set the number of channel paths to be $L = 3$, where one line-of-sight (LOS) path with the complex channel coefficient obeying $\lambda_1 \sim \mathcal{CN}(0, 1)$ and two non-line-of-sight (NLOS) paths with the complex channel coefficients obeying $\lambda_2, \lambda_3 \sim \mathcal{CN}(0, 0.01)$. For simplicity, we consider $N_1 = 2$.

As shown in Fig. 3, we compare the success rate for different beam training schemes. The successful rate is defined according to [13]. If the LOS path of the UE is correctly identified after beam training, we define that the beam training is successful; otherwise, we define that the beam training is failed. The ratio of the number of successful beam training over the total number of beam training is defined as the success rate. The curves shown in Fig. 3 are obtained by averaging the success rate over all K UEs. It is seen that the beam sweeping can achieve the best performance, as it always uses narrow beams with largest beam gain that can effectively combat the noise. The other three schemes have the similar performance. The TDMA-HBT is slightly better than our scheme and SMBT, because it always uses single-mainlobe beams that have better beam gain than the multi-mainlobe beams. Sine both SMBT and our scheme use the multi-mainlobe beams, their performance is almost the same.

As shown in Fig. 4, we compare the average sum-rate for different beam training schemes. The average sum-rate is

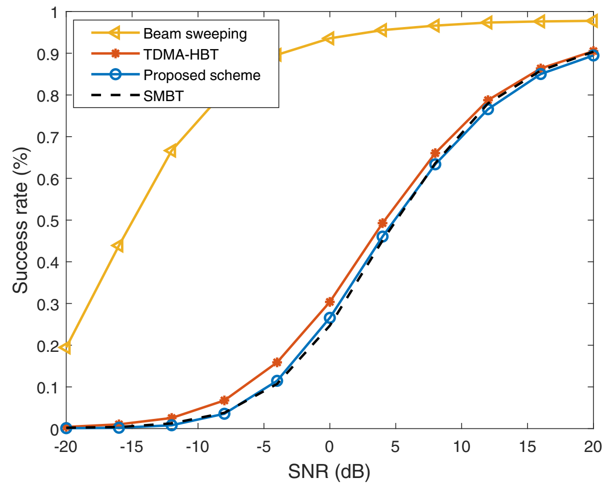

Fig. 3. Comparisons of success rate for different beam training schemes.

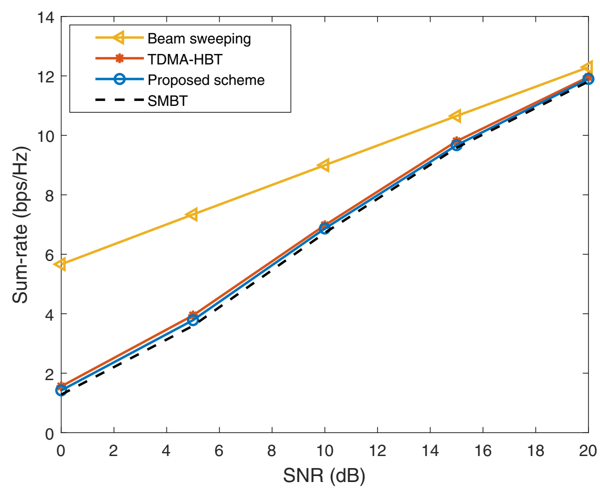

Fig. 4. Comparisons of the averaged sum-rate for different beam training schemes.

defined according to [13], where the achievable rate of all the UEs is summed together and then averaged over the number of UEs. It is seen that the beam sweeping performs the best and the other three schemes have the similar performance. As the SNR increases, the performance of TDMA-HBT, SMBT and our scheme approaches that of the beam sweeping. Therefore, our scheme has very similar performance as SMBT and TDMA-HBT, while 67% and 71% reduction in the number of feedbacks can be achieved, respectively.

V. Conclusions

In this paper, we have proposed a codeword design algorithm to design the hierarchical codebook. We have proposed a beam training scheme with limited feedback. The proposed beam training scheme includes two stages and requires only one feedback in each stage by using the designed hierarchical codebook. In the first stage, beam training for the top layer of our designed codebook is performed to narrow the search range of the channel angle of departure (AoD). In the second stage, based on the feedback information from the first stage, beam training for the other layers of our designed codebook is performed, where the BS can obtain an estimate of the channel AoD. Simulation results have shown that the performance of the proposed beam training scheme can approach that of TDMA hierarchical beam training scheme with much fewer feedbacks.

Acknowledgment

This work is supported in part by the National Natural Science Foundation of China (NSFC) under Grant 61871119 and by the Natural Science Foundation of Jiangsu Province under Grant BK20161428.

References

[1] R. W. Heath, N. Gonzalez-Prelcic, S. Rangan, W. Roh, and A. Sayeed, "An overview of signal processing techniques for millimeter wave MIMO systems," *IEEE J. Sel. Top. Signal Process.*, vol. 10, no. 3, pp. 436–453, Apr. 2016.

[2] O. E. Ayach, R. W. Heath, S. Abu-Surra, S. Rajagopal, and Z. Pi, "Low complexity precoding for large millimeter wave MIMO systems," in *Proc. IEEE ICC*, Ottawa, ON, Canada, June 2012, pp. 3724–3729.

[3] W. Ma, C. Qi, Z. Zhang, and J. Cheng, "Sparse channel estimation and hybrid precoding using deep learning for millimeter wave massive MIMO," *IEEE Trans. Commun.*, vol. 68, no. 5, pp. 2838–2849, May 2020.

[4] Z. Xiao, T. He, P. Xia, and X. G. Xia, "Hierarchical codebook design for beamforming training in millimeter-wave communication," *IEEE Trans. Wireless Commun.*, vol. 15, no. 5, pp. 3380–3392, May. 2016.

[5] J. Song, J. Choi, and D. J. Love, "Common codebook millimeter wave beam design: Designing beams for both sounding and communication with uniform planar arrays," *IEEE Trans. Commun.*, vol. 65, no. 4, pp. 1859–1872, Apr. 2017.

[6] A. Ali, N. Gonzálezz-Prelcic, and R. W. Heath, "Millimeter wave beam-selection using out-of-band spatial information," *IEEE Trans. Wireless Commun.*, vol. 17, no. 2, pp. 1038–1052, Feb. 2018.

[7] K. Chen and C. Qi, "Beam training based on dynamic hierarchical codebook for millimeter wave massive MIMO," *IEEE Commun. Lett.*, vol. 23, no. 1, pp. 132–135, Jan. 2019.

[8] R. Zhang, H. Zhang, W. Xu, and C. Zhao, "A codebook based simultaneous beam training for mmWave multi-user MIMO systems with split structures," in *Proc. 2018 IEEE Global Commun. Conf. (GLOBECOM)*, Abu Dhabi, UAE, Dec. 2018, pp. 1–6.

[9] K. Chen, C. Qi, O. A. Dobre, and G. Y. Li, "Simultaneous multiuser beam training using adaptive hierarchical codebook for mmWave massive MIMO," in *Proc. 2019 IEEE Global Commun. Conf. (GLOBECOM)*, Waikoloa Village, HI, USA, Dec. 2019, pp. 1–6.

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

[11] X. Sun, C. Qi, and G. Y. Li, "Beam training and allocation for multiuser millimeter wave massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 18, no. 2, pp. 1041–1053, Feb. 2019.

[12] A. Alkhateeb, G. Leus, and R. W. Heath, "Limited feedback hybrid precoding for multi-user millimeter wave systems," *IEEE Trans. Wireless Commun.*, vol. 14, no. 11, pp. 6481–6494, Nov. 2015.

[13] C. Qi, K. Chen, O. Dobre, and G. Y. Li, "Hierarchical codebook based multiuser beam training for millimeter massive MIMO," *IEEE Trans. Wireless Commun.*, vol. 19, no. 12, Dec. 2020.

[14] K. Chen, C. Qi, and G. Y. Li, "Two-step codeword design for millimeter wave massive MIMO systems with quantized phase shifters," *IEEE Trans. Signal Process.*, vol. 68, no. 1, pp. 170–180, Jan. 2020.

[15] H. He, P. Stoica, and J. Li, "Wideband MIMO systems: Signal design for transmit beampattern synthesis," *IEEE Trans. Signal Process.*, vol. 59, no. 2, pp. 618–628, Feb. 2011.