Hierarchical Codebook-Based Multiuser Beam Training for Millimeter Wave Massive MIMO

Chenhao Qi, *Senior Member, IEEE*, Kangjian Chen, *Member, IEEE*, Octavia A. Dobre, *Fellow, IEEE*, and Geoffrey Ye Li, *Fellow, IEEE*

Abstract—In this article, multiuser beam training based on hierarchical codebook for millimeter wave massive multi-input multi-output is investigated, where the base station (BS) simultaneously performs beam training with multiple user equipments (UEs). For the UEs, an alternative minimization method with a closed-form expression (AMCF) is proposed to design the hierarchical codebook under the constant modulus constraint. To speed up the convergence of the AMCF, an initialization method based on Zadoff-Chu sequence is proposed. For the BS, a simultaneous multiuser beam training scheme based on an adaptively designed hierarchical codebook is proposed, where the codewords in the current layer of the codebook are designed according to the beam training results of the previous layer. The codewords at the BS are designed with multiple mainlobes, each covering a spatial region for one or more UEs. Simulation results verify the effectiveness of the proposed hierarchical codebook design schemes and show that the proposed multiuser beam training scheme can approach the performance of the beam sweeping but with significantly reduced beam training overhead.

Index Terms—Beam training, hierarchical codebook, massive multi-input multi-output (MIMO), millimeter wave (mmWave) communications.

I. INTRODUCTION

MILLIMETER wave (mmWave) massive multi-input multi-output (MIMO) has been considered as a promising technology for future wireless communications due to its rich spectral resource [2]–[7]. However, the transmission of mmWave signal experiences large path loss because of its high frequency. To compensate it, antenna arrays with a hybrid precoding architecture have been introduced, where a small number of radio frequency (RF) chains are connected to a

large number of antennas via phase shifters [8]. Compared to the fully digital architecture, the hybrid precoding architecture can substantially reduce the hardware complexity and save the energy consumption.

To acquire the channel state information (CSI) in mmWave massive MIMO systems with the hybrid precoding structure, codebook-based beam training methods have been widely adopted [9]–[12]. To reduce the training overhead, hierarchical codebook-based beam training methods have been proposed. For hierarchical beam training, a predefined hierarchical codebook including several layers of codebooks is typically employed, where the spatial region covered by the codeword at an upper layer of the codebook is split into several smaller spatial regions covered by codewords at a lower layer [13]. Earlier work on hierarchical beam training focuses on peer-to-peer mmWave massive MIMO systems. In [14], a hierarchical codebook is utilized to acquire the CSI in mmWave massive MIMO systems, where channel estimation is formulated as a sparse reconstruction problem. For multiuser scenarios, a straightforward extension of the above work is time-division multiple access (TDMA) hierarchical beam training, where the base station (BS) sequentially performs the hierarchical beam training user by user and each occupies a different part of time. However, the total training overhead grows linearly with the number of users. To reduce the overhead of beam training, a simultaneous hierarchical beam training for a multiuser mmWave massive MIMO system is proposed for partially connected structure, where each RF chain is solely connected to an antenna subarray at the BS and the beam training is independently performed by each subarray [15].

So far, several methods have been proposed to design the hierarchical codebook [16]. Many methods exploit the degree of multiple RF chains to construct the codebook. If there is only one RF chain which is the general setup for most UEs, the following two methods can be employed to design the hierarchical codebook. In [13], a joint sub-array and de-activation (JOINT) method is proposed, where the weighted summation of several sub-arrays is used to form wide beams during codebook design. However, half of the antennas may be powered off for the JOINT method, which will weaken the signal strength and thus reduce the coverage. To address this issue, an enhanced JOINT (EJOINT) method is developed, where the codewords are formed without antenna de-activation [17].

In this article, we propose a simultaneous multiuser hierarchical beam training scheme based on our designed adaptive hierarchical codebook for multiuser mmWave massive

Fig. 1.   Illustration of a multiuser mmWave massive MIMO system with a BS and K UEs.

MIMO systems. The main contributions of this article are as follows.

1) For the UEs served by the BS, we propose an alternative minimization method with a closed-form expression (AMCF) to design the hierarchical codebook under the constant modulus constraint. To speed up the convergence of the AMCF, one initialization method is proposed, namely AMCF with Zadoff-Chu sequence initialization (AMCF-ZCI).

2) For the BS, rather than sequentially performing the beam training with different users using the same hierarchical codebook, we adaptively design the hierarchical codebook, where the codewords in the current layer of the hierarchical codebook are determined by the beam training results of the previous layer. In particular, we design multi-mainlobe codewords for the BS to simultaneously perform the beam training with all UEs, where each mainlobe covers a spatial region for one or more UEs. Excluding the bottom layer of the hierarchical codebook, there are only two codewords at each layer in the designed adaptive hierarchical codebook, which only requires two times of simultaneous beam training for all the UEs no matter how many UEs the BS serves. Compared with the existing beam training schemes, the proposed simultaneous multiuser hierarchical beam training scheme can substantially reduce the training overhead.

The rest of this article is organized as follows. The problem of multiuser beam training is formulated in Section II. The codebook design for the UEs is investigated in Section III. The codebook design for the BS and the simultaneous beam training is discussed in Section IV. Simulation results are provided in Section V. Finally, Section VI concludes this article.

The notations are defined as follows. Symbols for matrices (upper case) and vectors (lower case) are in boldface. $[\boldsymbol{a}]_n$, $[\boldsymbol{A}]_{:,n}$ and $[\boldsymbol{A}]_{m,n}$ denote the nth entry of a vector $\boldsymbol{a}$, the nth column of a matrix $\boldsymbol{A}$, and the entry on the mth row and nth column of $\boldsymbol{A}$. According to the convention, $\boldsymbol{I}$, $(\cdot)^T$, $(\cdot)^H$, $\|\cdot\|_F$,

$\|\cdot\|_2$, $\mathbb{C}$, $\mathbb{E}\{\cdot\}$, $\mathrm{diag}\{\cdot\}$, $\circ$ and $\mathcal{CN}$ denote the identity matrix, transpose, conjugate transpose (Hermitian), Frobenius norm, ℓ_2-norm, set of complex number, operation of expectation, diagonal matrix, entry-wise product and complex Gaussian distribution, respectively. $\angle(\cdot)$ denotes the phase of a complex value or a vector.

II. SYSTEM MODEL

As shown in Fig. 1, we consider a multiuser mmWave massive MIMO system with a BS and K UEs. The number of antennas at the BS and each UE is N_{BS} and $N_{\mathrm{UE}}(N_{\mathrm{UE}} \leq N_{\mathrm{BS}})$, respectively. The number of RF chains at the BS and each UE is $N_{\mathrm{RF}}(N_{\mathrm{RF}} \ll N_{\mathrm{BS}})$ and one, respectively. To simplify the analysis, both N_{BS} and N_{BS} are set as an integer power of two. The BS employs hybrid precoding, including digital precoding and analog precoding while each UE employs analog combining. At the BS, each RF chain is connected to N_{BS} antennas via N_{BS} quantized phase shifters. At each UE, one RF chain is connected to N_{UE} antennas via N_{UE} quantized phase shifters. The antennas at both the BS and the UEs are placed into uniform linear arrays (ULAs) with half wavelength spacing. Generally, each RF chain at the BS can support an independent data stream for a UE. Therefore the number of UEs simultaneously served by the BS is usually smaller than the number of RF chains, i.e., $K \leq N_{\mathrm{RF}}$.

During the downlink signal transmission from the BS to the UEs, the received signal by the kth UE for $k = 1, 2, \ldots, K$ can be expressed as

$$y_k = \boldsymbol{w}_k^H \boldsymbol{H}_k \boldsymbol{F}_{\mathrm{RF}} \boldsymbol{F}_{\mathrm{BB}} \boldsymbol{P} \boldsymbol{s} + \boldsymbol{w}_k^H \boldsymbol{n}_k, \qquad (1)$$

where y_k, $\boldsymbol{w}_k \in \mathbb{C}^{N_{\mathrm{UE}}}$, $\boldsymbol{H}_k \in \mathbb{C}^{N_{\mathrm{UE}} \times N_{\mathrm{BS}}}$, $\boldsymbol{F}_{\mathrm{RF}} \in \mathbb{C}^{N_{\mathrm{BS}} \times N_{\mathrm{RF}}}$, $\boldsymbol{F}_{\mathrm{BB}} \in \mathbb{C}^{N_{\mathrm{RF}} \times K}$, $\boldsymbol{P} \triangleq \mathrm{diag}\{\sqrt{P_1}, \sqrt{P_2}, \ldots, \sqrt{P_K}\}$, $\boldsymbol{s} \in \mathbb{C}^K$ and $\boldsymbol{n}_k \in \mathbb{C}^{N_{\mathrm{UE}}}$ denote the received signal, the analog combiner of the kth UE, the channel matrix between the BS and the kth UE, the analog precoder of the BS, the digital precoder of the BS, the diagonal power allocation matrix, the transmitted signal vector, and the additive white Gaussian noise (AWGN) vector obeying $\boldsymbol{n}_k \sim \mathcal{CN}(0, \sigma^2 \boldsymbol{I}_{N_{\mathrm{UE}}})$, respectively. Note that the hybrid

precoder, including the analog precoder and digital precoder, has no power gain, i.e., $\|\boldsymbol{F}_{\mathrm{RF}}\boldsymbol{F}_{\mathrm{BB}}\|_F^2 = K$. The power allocation subjects to the constraint that $\sum_{k=1}^{K} P_k = P_{\mathrm{Total}}$, where P_k denotes the power allocated to the kth UE for $k = 1, 2, \ldots, K$. Moreover, the transmitted signal vector $\boldsymbol{s}$ subjects to the unit power constraint that $\mathbb{E}\{\boldsymbol{s}\boldsymbol{s}^H\} = \boldsymbol{I}_K$.

According to the widely used Saleh-Valenzuela channel model [2], [18], the mmWave MIMO channel matrix, $\boldsymbol{H}_k \in \mathbb{C}^{N_{\mathrm{UE}} \times N_{\mathrm{BS}}}$, between the BS and the kth UE can be expressed as

$$\boldsymbol{H}_k = \sqrt{\frac{N_{\mathrm{BS}} N_{\mathrm{UE}}}{L_k}} \sum_{l=1}^{L_k} \lambda_l \boldsymbol{\alpha}(N_{\mathrm{UE}}, \theta_{\mathrm{UE}}^l) \boldsymbol{\alpha}^H(N_{\mathrm{BS}}, \theta_{\mathrm{BS}}^l) \quad (2)$$

where L_k, λ_l, θ_{UE}^l, and θ_{BS}^l denote the number of multi-path, the channel gain, the channel angle-of-arrival (AoA), and the channel angle-of-departure (AoD) of the lth path, respectively. In fact, $\theta_{\mathrm{UE}}^l = \cos(\omega_{\mathrm{UE}}^l)$ and $\theta_{\mathrm{BS}}^l = \cos(\omega_{\mathrm{BS}}^l)$, where ω_{UE}^l and ω_{BS}^l denote the physical AoA and AoD of the lth path, respectively. Since $\omega_{\mathrm{UE}}^l \in [0, 2\pi)$ and $\omega_{\mathrm{BS}}^l \in [0, 2\pi)$, we have $\theta_{\mathrm{UE}}^l \in [-1, 1]$ and $\theta_{\mathrm{BS}}^l \in [-1, 1]$. The channel steering vector in (2) is defined as

$$\boldsymbol{\alpha}(N, \theta) = \frac{1}{\sqrt{N}} \left[1, e^{j\pi\theta}, \cdots, e^{j(N-1)\pi\theta}\right]^T \quad (3)$$

where N is the number of antennas and θ is the channel AoA or AoD.

Our objective is to maximize the averaged sum-rate of K UEs by adjusting $\boldsymbol{F}_{\mathrm{RF}}$, $\boldsymbol{F}_{\mathrm{BB}}$, $\{\boldsymbol{w}_k\}_{k=1}^K$ and $\{P_k\}_{k=1}^K$. It can be expressed as the following optimization problem,

$$\max_{\substack{\boldsymbol{F}_{\mathrm{RF}}, \boldsymbol{F}_{\mathrm{BB}}, \\ \{\boldsymbol{w}_k, P_k\}_{k=1}^K}} \quad \sum_{k=1}^{K} \frac{1}{K} R_k \quad (4a)$$

$$\text{s.t.} \quad \left|[\boldsymbol{F}_{\mathrm{RF}}]_{n,t}\right| = \frac{1}{\sqrt{N_{\mathrm{BS}}}}, \left|[\boldsymbol{w}_k]_m\right| = \frac{1}{\sqrt{N_{\mathrm{UE}}}}, \quad (4b)$$

$$\left\|\boldsymbol{F}_{\mathrm{RF}}[\boldsymbol{F}_{\mathrm{BB}}]_{:,k}\right\|_2 = 1, \quad (4c)$$

$$P_1 + P_2 + \cdots + P_K = P_{\mathrm{Total}}, \quad (4d)$$

$$n = 1, 2, \ldots, N_{\mathrm{BS}}, \quad m = 1, 2, \ldots, N_{\mathrm{UE}}, \quad (4e)$$

$$t = 1, 2, \cdots, N_{\mathrm{RF}}, \quad k = 1, 2, \cdots K.$$

where

$$R_k = \log_2 \left(1 + \frac{P_k \left|\boldsymbol{w}_k^H \boldsymbol{H}_k \boldsymbol{F}_{\mathrm{RF}}[\boldsymbol{F}_{\mathrm{BB}}]_{:,k}\right|^2}{\sum_{i \neq k} P_i \left|\boldsymbol{w}_k^H \boldsymbol{H}_k \boldsymbol{F}_{\mathrm{RF}}[\boldsymbol{F}_{\mathrm{BB}}]_{:,i}\right|^2 + \sigma^2}\right) \quad (5)$$

is the achievable rate of the kth UE for $k = 1, 2, \ldots, K$. In the above, (4b) indicates that the entries in $\boldsymbol{F}_{\mathrm{RF}}$ and $\boldsymbol{w}_k$ satisfy the constant envelop constraint of phase shifters, (4c) indicates that the hybrid precoder does not provide power gain, and (4d) indicates the power allocation for different users.

In (5), $\boldsymbol{H}_k$ is required to calculate R_k. But the codebook-based beam training can avoid directly using $\boldsymbol{H}_k$, since the estimation of $\boldsymbol{H}_k$ incurs a large overhead. We denote the codebooks at the BS and each UE as $\mathcal{F} = \{\boldsymbol{f}_c^1, \boldsymbol{f}_c^2, \cdots, \boldsymbol{f}_c^{N_{\mathrm{BS}}}\}$ and $\mathcal{W} = \{\boldsymbol{w}_c^1, \boldsymbol{w}_c^2, \cdots, \boldsymbol{w}_c^{N_{\mathrm{UE}}}\}$, respectively, where

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

$$\boldsymbol{w}_c^m = \boldsymbol{\alpha}(N_{\mathrm{UE}}, -1 + (2m-1)/N_{\mathrm{UE}}). \quad (6)$$

The objective of beam training is to select K codewords from $\mathcal{F}$ for the BS and K codewords from $\mathcal{W}$ for K UEs to maximize the averaged sum-rate of the K UEs. Then (4) can be rewritten as

$$\max_{\substack{\boldsymbol{F}_{\mathrm{RF}}, \boldsymbol{F}_{\mathrm{BB}}, \\ \{\boldsymbol{w}_k, P_k\}_{k=1}^K}} \quad \sum_{k=1}^{K} \frac{1}{K} R_k \quad (7a)$$

$$\text{s.t.} \quad \boldsymbol{f}_k \in \mathcal{F}, \boldsymbol{w}_k \in \mathcal{W}, \quad \left\|\boldsymbol{F}_{\mathrm{RF}}[\boldsymbol{F}_{\mathrm{BB}}]_{:,k}\right\|_2 = 1, \quad (7b)$$

$$P_1 + P_2 + \cdots + P_K = P_{\mathrm{Total}}, \quad (7c)$$

$$k = 1, 2, \cdots, K$$

where $\boldsymbol{f}_k \triangleq [\boldsymbol{F}_{\mathrm{RF}}]_{:,k}$. In (7), the coupling among $\boldsymbol{F}_{\mathrm{RF}}$, $\boldsymbol{F}_{\mathrm{BB}}$ and $\{P_k\}_{k=1}^K$ makes it difficult to solve the problem directly. A typical method in the existing work is to design $\boldsymbol{w}_k$ and $\boldsymbol{F}_{\mathrm{RF}}$ first without considering $\boldsymbol{F}_{\mathrm{BB}}$ and $\{P_k\}_{k=1}^K$, and then design $\boldsymbol{F}_{\mathrm{BB}}$ under the power constraint of (4c) to eliminate multiuser interference [12]. Once the beam training is finished and the CSI is obtained, we design $\{P_k\}_{k=1}^K$ for data transmission, e.g., using the water-filling method. During the beam training, we do not consider the power allocation and use the equal power for different users.

Since designing $\boldsymbol{F}_{\mathrm{RF}}$ is essentially to find $\boldsymbol{f}_k$, our objective turns to find a pair of $\boldsymbol{f}_k$ and $\boldsymbol{w}_k$ best fit for $\boldsymbol{H}_k$, which can be expressed as

$$\max_{\boldsymbol{f}_k, \boldsymbol{w}_k} \quad \left|\boldsymbol{w}_k^H \boldsymbol{H}_k \boldsymbol{f}_k\right|$$

$$\text{s.t.} \quad \boldsymbol{f}_k \in \mathcal{F}, \quad \boldsymbol{w}_k \in \mathcal{W}. \quad (8)$$

A straightforward method to solve (8) is the exhaustive beam training, which is also called beam sweeping. It tests all possible pairs of $\boldsymbol{f}_k$ and $\boldsymbol{w}_k$ to find the best one. However, such a method takes a long time and therefore with a large overhead. If we denote the period of each test of a pair of $\boldsymbol{f}_k$ and $\boldsymbol{w}_k$ as a time slot, the exhaustive beam training needs totally $N_{\mathrm{BS}} N_{\mathrm{UE}}$ time slots. Note that the number of total time slots is independent of K since the UEs can simultaneously test the power of their received signal and eventually feed back the indices of the best codewords to the BS.

To reduce the overhead of exhaustive beam training, hierarchical beam training that is based on hierarchical codebooks, is widely adopted [13]. The hierarchical codebook typically consists of a small number of low-resolution codewords covering a wide angle at the upper layer of the codebook and a large number of high-resolution codewords offering high directional gain at the lower layer of the codebook. The hierarchical beam training usually first tests the mmWave channel with some low-resolution codewords at the upper layer and then narrows down the beam width layer by layer until a codeword pair at bottom layer is obtained. We denote the hierarchical codebooks employed at the BS and UEs as $\mathcal{V}_{\mathrm{BS}}$ and $\mathcal{V}_{\mathrm{UE}}$, respectively. The mth codeword at the sth layer of $\mathcal{V}_{\mathrm{UE}}$ for $s = 1, \cdots, T$ and $m = 1, 2, \cdots 2^s$ is denoted as $\mathcal{V}_{\mathrm{UE}}(s, m)$ where

$$T = \log_2 N_{\mathrm{UE}} \quad (9)$$

is the number of layers of $\mathcal{V}_{\mathrm{UE}}$. Note that the codewords at the bottom layer of $\mathcal{V}_{\mathrm{UE}}$ are exactly the same as the codewords

in $\mathcal{W}$, which implies that the motivation of hierarchical beam training is to adopt the merits of binary tree to improve the efficiency of beam training. Compared with the exhaustive beam training, the hierarchical beam training sequentially performed user by user in the TDMA fashion can reduce the overhead from $N_{BS}N_{UE}$ to $2K(\log_2 N_{BS} + \log_2 N_{UE})$. For example, if $K = 4$, $N_{BS} = 64$ and $N_{UE} = 16$, the hierarchical beam training can reduce the training overhead by 92.2% compared to the exhaustive beam training.

III. HIERARCHICAL CODEBOOK DESIGN FOR UEs

In this section, we design the hierarchical codebook for the UEs, while the hierarchical codebook designed for the BS will be addressed in the next section together with the beam training.

A. Codebook Design Based on Alternative Minimization Method

Since each UE normally has only one RF chain, the N_{UE} antennas at each UE are connected to the RF chain via phase shifters without digital precoding. Therefore, each entry of the codeword in the hierarchical codebook must be designed under the constant modulus constraint. In the following we propose an alternative minimization method with a closed-form expression (AMCF).

Denote the absolute beam gain of a codeword $\boldsymbol{v} \in \mathbb{C}^{N_{UE}}$ with beam coverage $\mathcal{I}_v = [\Omega_0, \Omega_0 + B]$ as $g(\Omega)$ with $\Omega \in [-1, 1]$, where $g(\Omega)$ is predefined as [9]

$$g(\Omega) = \begin{cases} \sqrt{2/B}, & \Omega \in \mathcal{I}_v, \\ 0, & \Omega \notin \mathcal{I}_v. \end{cases} \tag{10}$$

Then, we can formulate the codeword design problem as

$$\min_{\boldsymbol{v}} \frac{1}{2} \int_{-1}^{1} \left(g(\Omega) - \left| \boldsymbol{\alpha}(N_{UE}, \Omega)^H \boldsymbol{v} \right| \right)^2 d\Omega \tag{11a}$$

$$\text{s.t.} \ \left| [\boldsymbol{v}]_n \right|^2 = \frac{1}{N_{UE}}, \quad n = 1, 2, \cdots, N_{UE} \tag{11b}$$

where (11a) aims to minimize the mean-squared error between the predefined beam gain of the codeword and the practical beam gain of $\boldsymbol{v}$, and (11b) is the constant modulus constraint imposed by the phase shifters of the UEs.

We equally quantize the continuous channel AoA $[-1, 1]$ into $Q(Q > N_{UE})$ angles, where the qth angle is denoted as

$$\Omega_q = -1 + (2q - 1)/Q. \tag{12}$$

for $q = 1, 2, \cdots Q$. Denote

$$\boldsymbol{A} \triangleq \sqrt{N_{UE}}[\boldsymbol{\alpha}(N_{UE}, \Omega_1), \boldsymbol{\alpha}(N_{UE}, \Omega_2), \ldots, \boldsymbol{\alpha}(N_{UE}, \Omega_Q)] \tag{13}$$

where $\boldsymbol{A} \in \mathbb{C}^{N_{UE} \times Q}$ is a matrix made up of Q channel steering vectors. Define a vector $\boldsymbol{g} \in \mathbb{C}^Q$ with

$$[\boldsymbol{g}]_q = g(\Omega_q), \quad q = 1, 2, \cdots, Q \tag{14}$$

to represent the values of the predefined beam gains along the quantized angles. Then the continuous problem in (11) is converted into the discrete problem as

$$\min_{\boldsymbol{v}} \left\| \boldsymbol{g} - |\boldsymbol{A}^H \boldsymbol{v}| \right\|_2^2 \tag{15a}$$

$$\text{s.t.} \ \left| [\boldsymbol{v}]_n \right|^2 = \frac{1}{N_{UE}}, \quad n = 1, 2, \cdots, N_{UE}. \tag{15b}$$

As Q grows to be infinity, the solution of (15) will approach the solution of (11). By introducing a phase vector $\boldsymbol{\Theta} \in \mathbb{R}^Q$, we can further rewrite (15) as

$$\min_{\boldsymbol{v}, \boldsymbol{\Theta}} \left\| \boldsymbol{r} - \boldsymbol{A}^H \boldsymbol{v} \right\|_2^2 \tag{16a}$$

$$\text{s.t.} \ \left| [\boldsymbol{v}]_n \right|^2 = \frac{1}{N_{UE}}, \quad n = 1, 2, \cdots, N_{UE}. \tag{16b}$$

where

$$\boldsymbol{r} \triangleq \boldsymbol{g} \circ e^{j\boldsymbol{\Theta}} \tag{17}$$

with $[\boldsymbol{r}]_q = [\boldsymbol{g}]_q e^{j[\boldsymbol{\Theta}]_q}$ for $q = 1, 2, \ldots, Q$. Note that (15) and (16) have the same optimal solution of $\boldsymbol{v}$ because we can always design $\boldsymbol{\Theta} = \angle(\boldsymbol{A}^H \boldsymbol{v})$ so that $\boldsymbol{r}$ and $(\boldsymbol{A}^H \boldsymbol{v})$ have the same phase.

The problem in (16) can be solved by the alternative minimization method. We determine $\boldsymbol{\Theta}$ with fixed $\boldsymbol{v}$ and then determine $\boldsymbol{v}$ with fixed $\boldsymbol{\Theta}$, which is repeatedly executed until the maximum number of iterations is reached.

When determining $\boldsymbol{\Theta}$ with fixed $\boldsymbol{v}$, the optimal solution of $\boldsymbol{\Theta}$ can be written as

$$\boldsymbol{\Theta} = \angle(\boldsymbol{A}^H \boldsymbol{v}). \tag{18}$$

When determining $\boldsymbol{v}$ with fixed $\boldsymbol{\Theta}$, (16) can be written as

$$\min_{\boldsymbol{v}} \left\| \boldsymbol{r} - \boldsymbol{A}^H \boldsymbol{v} \right\|_2^2$$

$$\text{s.t.} \ \left| [\boldsymbol{v}]_n \right|^2 = \frac{1}{N_{UE}}, \quad n = 1, 2, \cdots, N_{UE}, \tag{19}$$

which has already been investigated [19], [20]. In [19], the problem is resolved by a successive closed-form (SCF) algorithm, which involves solving a series of convex equality constrained quadratic programs. In [20], it is resolved by Riemannian optimization algorithm. Different from [19] and [20], in this work we will show that the problem has a closed-form expression, which no longer needs running any algorithms and therefore has very low computational complexity.

The objective function in (19) can be written as

$$\begin{aligned} \left\| \boldsymbol{r} - \boldsymbol{A}^H \boldsymbol{v} \right\|_2^2 &= (\boldsymbol{r} - \boldsymbol{A}^H \boldsymbol{v})^H (\boldsymbol{r} - \boldsymbol{A}^H \boldsymbol{v}) \\ &= \boldsymbol{r}^H \boldsymbol{r} - \boldsymbol{r}^H \boldsymbol{A}^H \boldsymbol{v} - \boldsymbol{v}^H \boldsymbol{A} \boldsymbol{r} + \boldsymbol{v}^H \boldsymbol{A} \boldsymbol{A}^H \boldsymbol{v} \\ &= C - \boldsymbol{p}^H \boldsymbol{v} - \boldsymbol{v}^H \boldsymbol{p} \end{aligned} \tag{20}$$

where $C \triangleq \boldsymbol{r}^H \boldsymbol{r} + \boldsymbol{v}^H \boldsymbol{A} \boldsymbol{A}^H \boldsymbol{v}$ and $\boldsymbol{p} \triangleq \boldsymbol{A} \boldsymbol{r}$. Note that $\boldsymbol{r}^H \boldsymbol{r} = \boldsymbol{g}^H \boldsymbol{g}$ is a constant determined by the predefined absolute beam gain in (10). In addition, we have $\boldsymbol{A} \boldsymbol{A}^H = Q \boldsymbol{I}_{N_{UE}}$ because the entry on the mth row and the nth column of $\boldsymbol{A} \boldsymbol{A}^H$, for $m = 1, 2, \cdots N_{UE}$ and $n = 1, 2, \cdots N_{UE}$, can be expressed as

$$\begin{aligned} [\boldsymbol{A} \boldsymbol{A}^H]_{m,n} &= [\boldsymbol{A}]_{m,:}[\boldsymbol{A}^H]_{:,n} = \sum_{q=1}^{Q} e^{j(m-n)\pi\Omega_q} \\ &= \sum_{q=1}^{Q} e^{j(m-n)\pi(-1 + \frac{2q-1}{Q})} \\ &= e^{j(m-n)\pi(-1 - \frac{1}{Q})} \sum_{q=1}^{Q} e^{j(m-n)\pi\frac{2q}{Q}} \end{aligned}$$

$$= \begin{cases} Q, & m = n, \\ 0, & \text{others.} \end{cases} \tag{21}$$

Then we have $\boldsymbol{v}^H \boldsymbol{A} \boldsymbol{A}^H \boldsymbol{v} = Q \boldsymbol{v}^H \boldsymbol{v}$, which is also a constant due to the constant modulus constraint of $\boldsymbol{v}$. Therefore, C is a constant and (19) can be convert into

$$\max_{\boldsymbol{v}} \ \boldsymbol{p}^H \boldsymbol{v} + \boldsymbol{v}^H \boldsymbol{p}$$
$$\text{s.t. } \left| [\boldsymbol{v}]_n \right|^2 = \frac{1}{N_{\text{UE}}}, \quad n = 1, 2, \cdots, N_{\text{UE}}. \tag{22}$$

Obviously, (22) is equivalent to the following problem

$$\max_{\boldsymbol{u}} \ \boldsymbol{t}^T \boldsymbol{u}$$
$$\text{s.t. } [\boldsymbol{u}]_n^2 + [\boldsymbol{u}]_{n+N_{\text{UE}}}^2 = \frac{1}{N_{\text{UE}}}, \quad n = 1, 2, \cdots, N_{\text{UE}} \tag{23}$$

where

$$\boldsymbol{t} = \begin{bmatrix} \text{Re}\{\boldsymbol{p}\} \\ \text{Im}\{\boldsymbol{p}\} \end{bmatrix}, \quad \boldsymbol{u} = \begin{bmatrix} \text{Re}\{\boldsymbol{v}\} \\ \text{Im}\{\boldsymbol{v}\} \end{bmatrix}. \tag{24}$$

Note that (23) can be divided into N_{UE} mutually independent subproblems, where the nth subproblem for $n = 1, 2, \cdots, N_{\text{UE}}$ can be written as

$$\max_{[\boldsymbol{u}]_n, [\boldsymbol{u}]_{n+N_{\text{UE}}}} [\boldsymbol{t}]_n [\boldsymbol{u}]_n + [\boldsymbol{t}]_{n+N_{\text{UE}}} [\boldsymbol{u}]_{n+N_{\text{UE}}}$$
$$\text{s.t. } [\boldsymbol{u}]_n^2 + [\boldsymbol{u}]_{n+N_{\text{UE}}}^2 = \frac{1}{N_{\text{UE}}}, \quad n = 1, 2, \cdots, N_{\text{UE}}. \tag{25}$$

The optimal solution of (25) can be easily computed as

$$[\boldsymbol{u}]_n = \frac{[\boldsymbol{t}]_n}{\sqrt{N_{\text{UE}}([\boldsymbol{t}]_n^2 + [\boldsymbol{t}]_{n+N_{\text{UE}}}^2)}},$$
$$[\boldsymbol{u}]_{n+N_{\text{UE}}} = \frac{[\boldsymbol{t}]_{n+N_{\text{UE}}}}{\sqrt{N_{\text{UE}}([\boldsymbol{t}]_n^2 + [\boldsymbol{t}]_{n+N_{\text{UE}}}^2)}}. \tag{26}$$

According to (24), we have

$$[\boldsymbol{v}]_n = [\boldsymbol{u}]_n + j[\boldsymbol{u}]_{n+N_{\text{UE}}}, \quad n = 1, 2, \cdots, N_{\text{UE}} \tag{27}$$

which is the closed-form expression for (19). Therefore, we can determine $\boldsymbol{v}$ with fixed $\boldsymbol{\Theta}$.

We alternatively optimize $\boldsymbol{v}$ and $\boldsymbol{\Theta}$ until a predefined maximum number of iterations M is reached. To speed up the convergence of the AMCF, we consider the following two different initializations.

B. Zadoff-Chu Sequence Initialization

According to [21], we can initialize the nth entry of $\boldsymbol{v}$ as

$$[\boldsymbol{v}^{(0)}]_n = \begin{cases} \frac{1}{\sqrt{N_{\text{UE}}}} e^{j\pi(\frac{Bn^2}{2N_{\text{UE}}} + n\Omega_0)}, & N_{\text{UE}} \text{ is even} \\ \frac{1}{\sqrt{N_{\text{UE}}}} e^{j\pi(\frac{Bn(n+1)}{2N_{\text{UE}}} + n\Omega_0)}, & N_{\text{UE}} \text{ is odd} \end{cases} \tag{28}$$

for $n = 1, 2, \ldots, N_{\text{UE}}$, where $\boldsymbol{v}^{(0)}$ is essentially a variant of Zadoff-Chu sequence in [22].

Although the beam generated by $\boldsymbol{v}^{(0)}$ covers the angle space of $[\Omega_0, \Omega_0 + B]$, $\boldsymbol{v}^{(0)}$ cannot be taken as a good codeword for

---

Algorithm 1 AMCF-ZCI Codeword Design for the UE

1: **Input:** N_{UE}, Ω_0, B and M.
2: Obtain $\boldsymbol{v}^{(0)}$ via (28).
3: Set $m = 1$.
4: **while** $m \leq M$ **do**
5:    Obtain $\boldsymbol{\Theta}^{(m)} = \angle(\boldsymbol{A}^H \boldsymbol{v}^{(m-1)})$ according to (18).
6:    Obtain $\boldsymbol{r}^{(m)} = \boldsymbol{g} \circ e^{j\boldsymbol{\Theta}^{(m)}}$ according to (17).
7:    Obtain $\boldsymbol{v}^{(m)}$ via (27).
8:    $m \leftarrow m + 1$.
9: **end while**
10: **Output:** $\boldsymbol{v}_{\text{o}} = \boldsymbol{v}^{(M)}$.

---

beam training. To be specific, in the transition zone of the beam generated by $\boldsymbol{v}^{(0)}$, there are lots of fluctuations, which can deteriorate the performance of beam training. To tackle this issue, additional hardware such as a group of unequal power dividers, is used in the phase shifter network so that the power of each antenna can be changed to smooth the transition zone [21].

Different from [21] that uses a group of unequal power dividers, in this work we avoid any additional hardware. We just use $\boldsymbol{v}^{(0)}$ as the initialization of AMCF and use the output of AMCF $\boldsymbol{v}_{\text{o}}$ as a designed codeword for the UE.

The steps of the AMCF with Zadoff-Chu sequence initialization (AMCF-ZCI) are summarized in **Algorithm 1**.

Finally, based on the aforementioned codeword design schemes, the hierarchical codebook for each UE can be designed as follows.

1) Initialize the layer counter of the codebook as $s = 1$ and the left boundary of the beam coverage as $\Omega_0 = -1$. Then set $B = 2/2^s$.
2) Design $\boldsymbol{\mathcal{V}}_{\text{UE}}(s, 1)$ by **Algorithm 1**.
3) Obtain $\boldsymbol{\mathcal{V}}_{\text{UE}}(s, m)$, for $m = 2, \cdots 2^s$, based on the following equation

$$\boldsymbol{\mathcal{V}}_{\text{UE}}(s, m) = \sqrt{N_{\text{UE}}} \boldsymbol{\mathcal{V}}_{\text{UE}}(s, 1)$$
$$\circ \boldsymbol{\alpha}(N_{\text{UE}}, (m-1)/2^{s-1}) \tag{29}$$

which is essentially the shifted version of $\boldsymbol{\mathcal{V}}_{\text{UE}}(s, 1)$ for different beam coverage.

4) Increase s by one, i.e., $s \leftarrow s + 1$.
5) Repeat the steps 2) to 4) until reaching the last layer of $\boldsymbol{\mathcal{V}}_{\text{UE}}$, i.e., $s = T + 1$.

IV. SIMULTANEOUS BEAM TRAINING BASED ON ADAPTIVE HIERARCHICAL CODEBOOK

In this section, we will propose a simultaneous multi-user beam training scheme based on adaptive hierarchical codebook, which considerably reduces the training overhead compared to the existing hierarchical beam training.

A. Simultaneous Multiuser Hierarchical Beam Training

We denote the hierarchical codebook for the BS as $\boldsymbol{\mathcal{C}}$ to distinguish with the existing hierarchical codebook $\boldsymbol{\mathcal{V}}_{\text{BS}}$.
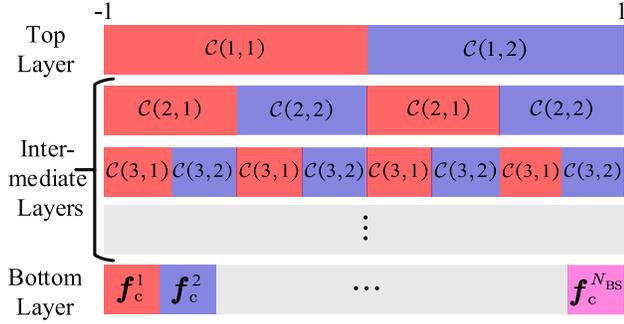
Fig. 2. Illustration of adaptive hierarchical codebook $\mathcal{C}$.

As shown in Fig. 2, the adaptive hierarchical codebook with totally $S = \log_2 N_{\text{BS}}$ layers can be divided into the top layer, the bottom layer and the intermediate layers.

1) In the top layer of the codebook, we equally divide the channel AoD $[-1, 1]$ into two codewords, so that the beam width of each codeword is one.

2) The bottom layer that is the Sth layer of the hierarchical codebook, is exactly the same as the bottom layer of the existing hierarchical codebook and can be designed according to (6).

3) The intermediate layers consist of the second to the $(S-1)$th layers of the codebook. Different from the existing codebook, each intermediate layer only includes two codewords, no matter how large K is. The beam coverage of codewords in the intermediate layers is intermittent. In particular, the codewords are adaptively designed according to the estimated channel AoDs in the previous layer. Note that the summation of the beam coverage of two codewords in the same layer may not be $[-1, 1]$ because some regions may not have any channel path and we do not need to waste signal beam to cover them.

Now we focus on designing codewords in the first and intermediate layers of the adaptive hierarchical codebook. In general, the beam coverage of a codeword at the upper layer can be considered as the union of several codewords at the bottom layer. Therefore, we can design the codewords in the first and intermediate layers by combining several codewords from the bottom layer of $\mathcal{C}$. Then the mth codeword at the sth layer of $\mathcal{C}$, denoted as $\mathcal{C}(s, m)$, for $s = 1, \cdots, S - 1$ and $m = 1, 2, \cdots 2^s$, can be represented as

$$\mathcal{C}(s, m) = \sum_{n \in \mathbf{\Psi}_{s,m}} e^{j\psi_n} \boldsymbol{f}_c^n \qquad (30)$$

which is essentially a weighted summation of several channel steering vectors. The indices of the codewords of $\mathcal{F}$ involved in the weighted summation form an integer set $\mathbf{\Psi}_{s,m}$. Here we introduce phase ψ_n to explore the additional degree of freedom to avoid low beam gain within the beam coverage [23]. Based on our previous work [24], [25], we can set

$$\psi_n = n\pi \left(-1 + \frac{1}{N_{\text{BS}}} \right). \qquad (31)$$

To fairly compare different codewords in each test, we usually normalize $\mathcal{C}(s, m)$ so that $\|\mathcal{C}(s, m)\|_2 = 1$.

We design $\mathbf{\Psi}_{s,m}$, for $s = 1, \cdots, S-1$ and $m = 1, 2, \cdots 2^s$, as follows. We denote the beam coverage of $\mathcal{C}(s, m)$ as $B_{s,m}$. Then $B_{1,m}$ at the top layer can be expressed as

$$B_{1,m} = [m - 2, m - 1], \quad m = 1, 2. \qquad (32)$$

We determine $\mathbf{\Psi}_{1,m}$ by

$$\mathbf{\Psi}_{1,m} = \left\{ n \left| \left[-1 + \frac{2n-2}{N_{\text{BS}}}, -1 + \frac{2n}{N_{\text{BS}}} \right] \subseteq B_{1,m}, \right. \right.$$
$$\left. n = 1, 2, \cdots, N_{\text{BS}} \right\} \qquad (33)$$

for $m = 1, 2$, where $[-1 + \frac{2n-2}{N_{\text{BS}}}, -1 + \frac{2n}{N_{\text{BS}}}]$ denotes the beam coverage of $\boldsymbol{f}_c^n$.

The BS sequentially transmits $\mathcal{C}(1, 1)$ and $\mathcal{C}(1, 2)$ to all K UEs and each UE receives the signal with $\mathcal{V}_{\text{UE}}(1, 1)$ and $\mathcal{V}_{\text{UE}}(1, 2)$, respectively. Then each UE compares the received signal power of $\mathcal{C}(1, 1)$ and $\mathcal{C}(1, 2)$ and individually feeds back the index of the larger codeword to the BS. Denote $\mathcal{K} \triangleq \{1, 2, \ldots, K\}$. We define a vector $\mathbf{\Gamma}_1$ of length of K, where the $k(k \in \mathcal{K})$th entry of $\mathbf{\Gamma}_1$ denoted as $[\mathbf{\Gamma}_1]_k$ corresponds to the index of the larger codewords from the kth UE, i.e., $[\mathbf{\Gamma}_1]_k \in \{1, 2\}$.

Denote the index set of the selected codewords after beam training at the $(s-1)$th layer to be $\mathbf{\Gamma}_{s-1}$ for $s = 2, 3, \cdots, S - 1$. According to the existing hierarchical beam training, at the sth layer, the BS will test $\mathcal{V}_{\text{BS}}(s, 2[\mathbf{\Gamma}_{s-1}]_k - 1)$ and $\mathcal{V}_{\text{BS}}(s, 2[\mathbf{\Gamma}_{s-1}]_k)$, which are the refined codewords of $\mathcal{V}_{\text{BS}}(s - 1, [\mathbf{\Gamma}_{s-1}]_k)$. Therefore, it needs totally $2K$ times of beam training to test all K UEs. To reduce the training overhead, we consider the following two cases:

1) If $[\mathbf{\Gamma}_{s-1}]_i = [\mathbf{\Gamma}_{s-1}]_q$ $(i \in \mathcal{K}, q \in \mathcal{K}, i \neq q)$ that is the ith UE and the qth UE share the same AoD at the $(s-1)$th layer of $\mathcal{V}_{\text{BS}}$, we can perform beam training for them simultaneously because they will have the same refined codewords at the sth layer of $\mathcal{V}_{\text{BS}}$.

2) If $[\mathbf{\Gamma}_{s-1}]_i \neq [\mathbf{\Gamma}_{s-1}]_q$ $(i \in \mathcal{K}, q \in \mathcal{K}, i \neq q)$, which means that the ith UE and the qth UE have different AoD at the $(s-1)$th layer of $\mathcal{V}_{\text{BS}}$, the ith UE cannot receive the signal transmitted from the beam coverage of $\mathcal{V}_{\text{BS}}(s - 1, [\mathbf{\Gamma}_{s-1}]_q)$ because the AoD of the ith UE is located in the beam coverage of $\mathcal{V}_{\text{BS}}(s - 1, [\mathbf{\Gamma}_{s-1}]_i)$. Therefore, we can distinguish different UEs based on their different AoDs. We will show in the following texts that the BS can also simultaneously perform the beam training for these two UEs.

Based on the above discussion, in either case, we suppose $\mathbf{\Gamma}_{s-1}$ has $K'(K' \leq K)$ different integers, which corresponds to K' different codewords at the $(s - 1)$th layer and $2K'$ refined codewords at the sth layer of $\mathcal{V}_{\text{BS}}$. In the proposed simultaneous multiuser beam training scheme, we divide these $2K'$ refined codewords into two groups. The union of beam coverage of K' codewords in the first and second group are

respectively denoted as

$$
\begin{cases}
B_{s,1} = \bigcup_m D_{s,m}, & \text{if } \dfrac{m+1}{2} \in \Gamma_{s-1}, \ m = 1, 2, \cdots, 2^s, \\
B_{s,2} = \bigcup_m D_{s,m}, & \text{if } \dfrac{m}{2} \in \Gamma_{s-1}, \ m = 1, 2, \cdots, 2^s
\end{cases}
$$

$$(34)$$

where $D_{s,m} = [-1+(m-1)/2^{s-1}, -1+m/2^{s-1}]$ is the beam coverage of $\mathcal{V}_{\text{BS}}(s, m)$. It is seen that the beam coverage of $B_{s,1}$ and $B_{s,2}$ is intermittent, which is the motivation of our work to design multi-mainlobe codewords. Then we can obtain $\Psi_{s,1}$ and $\Psi_{s,2}$ based on $B_{s,1}$ and $B_{s,2}$ respectively via

$$
\Psi_{s,m} = \left\{ n \,\middle|\, \left[-1 + \frac{2n-2}{N_{\text{BS}}}, -1 + \frac{2n}{N_{\text{BS}}} \right] \subset B_{s,m}, \right.
$$
$$
\left. n = 1, 2, \cdots, N_{\text{BS}} \right\} \quad (35)
$$

for $m = 1, 2$. Given $\Psi_{s,1}$ and $\Psi_{s,2}$, we can design $\mathcal{C}(s, 1)$ and $\mathcal{C}(s, 2)$, respectively via (30). Therefore, by using (34), (35) and (30), we can design $\mathcal{C}(s, 1)$ and $\mathcal{C}(s, 2)$ based on the beam training results of the $(s-1)$th layer, i.e., Γ_{s-1}. Note that either $\mathcal{C}(s, 1)$ or $\mathcal{C}(s, 2)$ is a multi-mainlobe codeword, where each mainlobe covers a spatial region that one or more users are probably in. When the number of UEs increases, the number of mainlobes of the multi-mainlobe codeword may also grow; but the number of codewords keeps to be two in each layer excluding the bottom layer of the hierarchical codebook.

At the sth layer of $\mathcal{C}$, for $s = 2, 3, \ldots, S - 1$, the BS sequentially transmits $\mathcal{C}(s, 1)$ and $\mathcal{C}(s, 2)$ to all K UEs. Note that there are only two codewords at each intermediate layer, which only requires two times of simultaneous beam training for all the UEs no matter how many UEs the BS serves. Then each UE compares the received signal power of $\mathcal{C}(s, 1)$ and $\mathcal{C}(s, 2)$ and individually feeds back the index of the larger codeword to the BS. We define Φ_s as a vector of length of K to keep the indices fed back by all K UEs, where $[\Phi_s]_k$ is the index fed back by the kth UE. In fact, $[\Phi_s]_k \in \{1, 2\}$. Then we can obtain Γ_s by

$$
[\Gamma_s]_k = 2([\Gamma_{s-1}]_k - 1) + [\Phi_s]_k, \quad (36)
$$

which can be used to determine $B_{s+1,1}$ and $B_{s+1,2}$ via (34) for $k = 1, 2, \cdots, K$. We iteratively perform these steps until arriving at the bottom layer of $\mathcal{C}$.

At the bottom layer of $\mathcal{C}$, different from the downlink beam training in the top and intermediate layers, we perform the uplink beam training so that each entry of the effective channel matrix in (40) can be obtained. During the uplink beam training between the kth UE and the BS, the BS sequentially uses $f_c^{2[\Gamma_{S-1}]_k - 1}$ and $f_c^{2[\Gamma_{S-1}]_k}$ to receive the signal and selected one with the larger received signal power. In this way, the best BS codeword in $\mathcal{F}$ for the kth UE can be determined. Note that the BS has N_{RF} RF chains, which implies that the BS can use multiple RF chains for parallel signal receiving to improve the efficiency [12], [26]. Therefore, totally $2K$ times of beam training are required at the bottom layer of $\mathcal{C}$. Similar to (36), we can obtain Γ_S by

$$
[\Gamma_S]_k = 2([\Gamma_{S-1}]_k - 1) + [\Phi_S]_k, \quad k = 1, 2, \cdots, K. \quad (37)
$$

Algorithm 2 Simultaneous Multiuser Hierarchical Beam Training

1: **Input:** N_{BS}, N_{UE} and K.
2: Obtain $\mathcal{C}(1, 1)$ and $\mathcal{C}(1, 2)$ via (33) and (30).
3: Obtain Γ_1 by the top layer beam training.
4: Set $S = \log_2 N_{\text{BS}}$.
5: **for** $s = 2, 3, \ldots, S - 1$ **do**
6:      Obtain $B_{s,1}$ and $B_{s,2}$ via (34).
7:      Obtain $\Psi_{s,1}$ and $\Psi_{s,2}$ via (35).
8:      Generate $\mathcal{C}(s, 1)$ and $\mathcal{C}(s, 2)$ via (30).
9:      Obtain Γ_s via (36).
10: **end for**
11: Obtain Γ_S via (37).
12: Obtain $\widehat{f}_k$ via (38).
13: **Output:** $\{\widehat{f}_k, k = 1, 2, \ldots, K\}$.
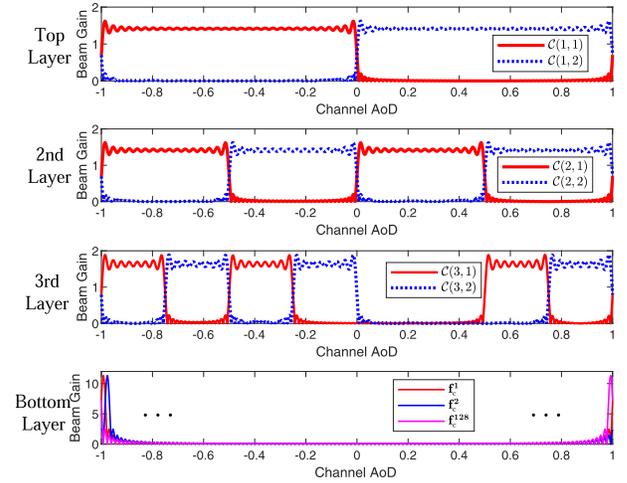
Fig. 3. Beam gain of different codewords in adaptive hierarchical codebook $\mathcal{C}$ with $N_{\text{BS}} = 128$, $N_{\text{UE}} = 16$ and $K = 4$.

Finally, the kth column of the designed analog precoder $\widehat{F}_{\text{RF}}$, denoted as $\widehat{f}_k$, can be obtained via

$$
\widehat{f}_k = f_c^{[\Gamma_S]_k}. \quad (38)
$$

The detailed steps of the proposed simultaneous multiuser hierarchical beam training is summarized in **Algorithm** 2.

When designing codewords in the top and intermediate layers of $\mathcal{C}$, we first obtain the ideal codewords by the weighted summation of channel steering vectors as (30) and then obtain the practical codewords regarding the number of RF chains and the resolution of phase shifters to approximate the ideal codewords based on the method in [25].

Now we give an example for the proposed simultaneous multiuser hierarchical beam training with $N_{\text{BS}} = 128$, $N_{\text{UE}} = 16$ and $K = 4$. As shown in Fig. 3, we illustrate the beam gain of different codewords in $\mathcal{C}$. To improve the readability of our scheme, each layer in Fig. 3 corresponds to that in Fig. 2. At the top layer of $\mathcal{C}$, the BS sequentially transmits $\mathcal{C}(1, 1)$ and $\mathcal{C}(1, 2)$. The union of beam coverage of $\mathcal{C}(1, 1)$ and $\mathcal{C}(1, 2)$ equals the full space of $[-1, 1]$, since the BS has no knowlege of the UEs. After the top layer beam training, we suppose

Schemes	Training Overhead	Feedback Overhead
Our scheme	$2(K + \log_2 N_{\mathrm{UE}} N_{\mathrm{BS}} - 1)$	$K(\log_2 N_{\mathrm{BS}} - 1)$
Scheme in [11]	$N_{\mathrm{BS}} N_{\mathrm{UE}}$	K
TDMA hierarchical beam training	$2K(\log_2 N_{\mathrm{BS}} + \log_2 N_{\mathrm{UE}})$	$K \log_2 N_{\mathrm{BS}}$

the indices fed back from the four UEs form a set $\boldsymbol{\Gamma}_1 = \{1, 1, 2, 2\}$, which indicates that the channel AoDs of the first and second UEs happen to locate in the same beam coverage of $\boldsymbol{\mathcal{C}}(1, 1)$, and the channel AoDs of the third and fourth UEs happen to locate in the same beam coverage of $\boldsymbol{\mathcal{C}}(1, 2)$. Based on $\boldsymbol{\Gamma}_1$, we can obtain $\boldsymbol{\Psi}_{2,1} = \{1, 2 \cdots, 32, 65, 66, \cdots, 96\}$ and $\boldsymbol{\Psi}_{2,2} = \{33, 34 \cdots, 64, 97, 98, \cdots, 128\}$ via (35). Based on $\boldsymbol{\Psi}_{2,1}$ and $\boldsymbol{\Psi}_{2,2}$, we can design two multi-mainlobe codewords $\boldsymbol{\mathcal{C}}(2, 1)$ and $\boldsymbol{\mathcal{C}}(2, 2)$ via (30). During the second layer of beam training, the BS sequentially transmits $\boldsymbol{\mathcal{C}}(2, 1)$ and $\boldsymbol{\mathcal{C}}(2, 2)$. Suppose the indices fed back from the four UEs form a set $\boldsymbol{\Phi}_2 = \{1, 2, 2, 2\}$, where each entry denotes the codeword index $\{i | \boldsymbol{\mathcal{C}}(2, i), i = 1, 2\}$ with the larger received signal power at the UEs. Then we can obtain $\boldsymbol{\Gamma}_2 = \{1, 2, 4, 4\}$ via (36). Based on $\boldsymbol{\Gamma}_2$, we can design $\boldsymbol{\Psi}_{3,1} = \{1, 2, \cdots, 16, 33, 34, \cdots, 48, 97, 98 \cdots, 112\}$ and $\boldsymbol{\Psi}_{3,2} = \{17, 18, \cdots, 32, 49, 50, \cdots, 64, 113, 114, \cdots, 128\}$ via (35). Based on $\boldsymbol{\Psi}_{3,1}$ and $\boldsymbol{\Psi}_{3,2}$, we can design two multi-mainlobe codewords $\boldsymbol{\mathcal{C}}(3, 1)$ and $\boldsymbol{\mathcal{C}}(3, 2)$ via (30). Note that both $\boldsymbol{\mathcal{C}}(3, 1)$ and $\boldsymbol{\mathcal{C}}(3, 2)$ have three mainlobes. We repeat these procedures until arriving at the bottom layer of $\boldsymbol{\mathcal{C}}$.

B. Digital Precoding

Note that the beam training and data transmission are two different stages of mmWave massive MIMO communications, where the former is to obtain the CSI that will be used for the latter. During the beam training, the BS employs a hierarchical codebook with multi-mainlobe codewords to serve all the users, where multiple RF chains might be used to generate multi-mainlobe codewords. Once the beam training is finished, the BS finds a best codeword $\widehat{\boldsymbol{f}}_k$ for the kth user, where $\widehat{\boldsymbol{f}}_k$ can be generated by a single RF chain according to (6). Since the designed analog combiner, $\widehat{\boldsymbol{w}}_k$, for the kth UE can be obtained by the existing hierarchical beam training method, the details are omitted in this work due to the page limitation.

In the following, we design the digital precoding for the data transmission. Stacking $\{y_k, k = 1, 2, \ldots, K\}$ in (1) together as $\boldsymbol{y} = [y_1, y_2, \ldots, y_K]^T$, we have

$$\boldsymbol{y} = \boldsymbol{H}_{\mathrm{e}} \boldsymbol{F}_{\mathrm{BB}} \boldsymbol{s}, \qquad (39)$$

where

$$\boldsymbol{H}_{\mathrm{e}} = \begin{bmatrix} \widehat{\boldsymbol{w}}_1^H \boldsymbol{H}_1 \widehat{\boldsymbol{f}}_1 & \widehat{\boldsymbol{w}}_1^H \boldsymbol{H}_1 \widehat{\boldsymbol{f}}_2 & \cdots & \widehat{\boldsymbol{w}}_1^H \boldsymbol{H}_1 \widehat{\boldsymbol{f}}_K \\ \widehat{\boldsymbol{w}}_2^H \boldsymbol{H}_2 \widehat{\boldsymbol{f}}_1 & \widehat{\boldsymbol{w}}_2^H \boldsymbol{H}_2 \widehat{\boldsymbol{f}}_2 & \cdots & \widehat{\boldsymbol{w}}_2^H \boldsymbol{H}_2 \widehat{\boldsymbol{f}}_K \\ \vdots & \cdots & \ddots & \vdots \\ \widehat{\boldsymbol{w}}_K^H \boldsymbol{H}_K \widehat{\boldsymbol{f}}_1 & \widehat{\boldsymbol{w}}_K^H \boldsymbol{H}_K \widehat{\boldsymbol{f}}_2 & \cdots & \widehat{\boldsymbol{w}}_K^H \boldsymbol{H}_K \widehat{\boldsymbol{f}}_K \end{bmatrix} \qquad (40)$$

is defined as the effective channel matrix. Note that each entry of $\boldsymbol{H}_{\mathrm{e}}$ can be obtained via the uplink beam training at the

bottom layer of $\boldsymbol{\mathcal{C}}$. If two users are geographically close to each other, they may share the same BS codeword, e.g., $\widehat{\boldsymbol{f}}_1 = \widehat{\boldsymbol{f}}_2$, which will cause the rank deficiency of $\boldsymbol{H}_{\mathrm{e}}$. Since the digital precoding requires $\boldsymbol{H}_{\mathrm{e}}$ to be full rank, we need to make beam allocation for different users to avoid beam conflict, which has already been addressed in [12] and is out of scope of this article.

The designed digital precoder under the zero forcing (ZF) criterion can be expressed as

$$\widehat{\boldsymbol{F}}_{\mathrm{BB}} = \boldsymbol{H}_{\mathrm{e}}^H (\boldsymbol{H}_{\mathrm{e}} \boldsymbol{H}_{\mathrm{e}}^H)^{-1}. \qquad (41)$$

C. Overhead Analysis

At the top layer of $\boldsymbol{\mathcal{C}}$, the BS sequentially transmits two codewords and each UE receives the signal with two codewords, which occupies 4 time slots. At the intermediate layers of $\boldsymbol{\mathcal{C}}$ from $s = 2$ to $s = \log_2 N_{\mathrm{UE}}$, the BS sequentially transmits two codewords and each UE receives signal with two codewords, which occupies totally $4(\log_2 N_{\mathrm{UE}} - 1)$ time slots. At the intermediate layers of $\boldsymbol{\mathcal{C}}$ from $s = \log_2 N_{\mathrm{UE}} + 1$ to $s = \log_2 N_{\mathrm{BS}} - 1$, where the hierarchical beam training at the UEs has already been finished, the BS transmits two codewords and each UE receives signal with a single codeword, which occupies totally $2(\log_2 N_{\mathrm{BS}} - \log_2 N_{\mathrm{UE}} - 1)$ time slots. At the bottom layer of $\boldsymbol{\mathcal{C}}$, the BS uses two codewords to receive the signal from each UE, which results in totally $2K$ time slots. In all, our proposed scheme needs totally $(2K + 2 \log_2 N_{\mathrm{UE}} N_{\mathrm{BS}} - 2)$ time slots. As shown in Table. I, we compare the training overhead of different schemes. For example, if $N_{\mathrm{BS}} = 128$, $N_{\mathrm{UE}} = 16$, $K = 8$, our scheme, the scheme in [11] and the TDMA hierarchical beam training require 36, 2048 and 176 time slots, respectively. Compared to the latter two schemes, our scheme can reduce the training overhead by 98.2% and 79.6%, respectively.

We also compare the feedback overhead from the UEs to the BS. For the scheme in [11], each UE needs only one time of feedback after finishing the beam sweeping, which results in totally K times of feedback. Since we do need feedback at the bottom layer in our scheme, the feedback times of our scheme is K less than that of the TDMA hierarchical beam training.

V. SIMULATION RESULTS

Now we evaluate our schemes for multiuser mmWave massive MIMO systems by simulation.

A. Evaluation of Codeword Design Schemes for Each UE

To evaluate the codeword design schemes for each UE, we consider a single-user mmWave massive MIMO system,
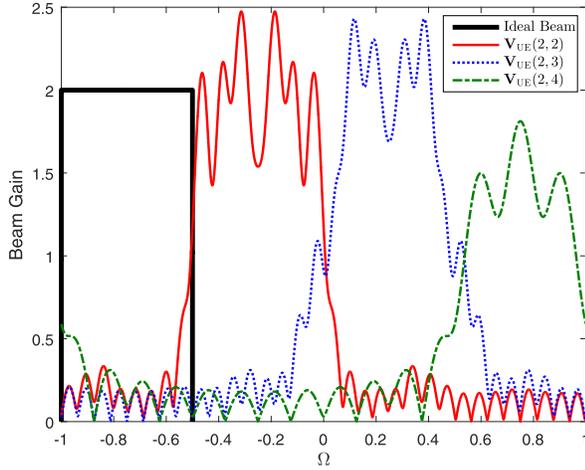
Fig. 4. Comparisons of beam patterns using different codeword design schemes for each UE.

Fig. 5. Comparisons of beam training performance in terms of success rate using the hierarchical codebooks designed by different schemes.

where the BS equipped with $N_{\text{BS}} = 128$ antennas serves only one UE for simplification, since there is no difference in codebook design between one UE and multiple UEs. The UE is supposed to be equipped with $N_{\text{UE}} = 32$ antennas and a single RF chain.

Fig. 4 compares the beam patterns for each UE using different codeword design schemes. For fair comparison, the beam width is set as $B = 1/2$, which is typically set for the codewords at the second layer of $\boldsymbol{V}_{\text{UE}}$. Given $N_{\text{UE}} = 32$, the constant modulus constraint on the antenna weights is $1/\sqrt{32}$. According to (10), the ideal beam gain is

$$g(\Omega) = \begin{cases} 2, & \Omega \in \mathcal{I}_v, \\ 0, & \Omega \notin \mathcal{I}_v, \end{cases} \qquad (42)$$

which can form the beam pattern illustrated by the black solid line in Fig. 4. The beam patterns of $\boldsymbol{\mathcal{V}}_{\text{UE}}(2,2)$, $\boldsymbol{\mathcal{V}}_{\text{UE}}(2,3)$ and $\boldsymbol{\mathcal{V}}_{\text{UE}}(2,4)$ are generated using the AMCF-ZCI, enhanced JOINT (EJOINT) and JOINT codeword design schemes, respectively. As a comparison, the ideal beam pattern of $\boldsymbol{\mathcal{V}}_{\text{UE}}(2,1)$ is also provided. From Fig. 4, it is seen that the performance of AMCF-ZCI outperforms that of JOINT and EJOINT. To be specific, EJOINT and JOINT have wider transition band than AMCF-ZCI, since the former two are based on the sub-array combining technique. In particular, the transition band of EJOINT is not monotonous, which may result in the failure of beam training. Moreover, the beam gain of JOINT is lower than AMCF-ZCI because half of antennas are closed for JOINT. Note that AMCF-ZCI can also design codewords with arbitrary beam width, which cannot be achieved by JOINT or EJOINT.

Now we compare the beam training performance in terms of success rate using the hierarchical codebooks designed by different schemes. The success rate is defined as follows. If the line-of-sight (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 BS uses the hierarchical codebook designed according to [24], while the UE uses the codebooks designed by AMCF-ZCI, EJOINT
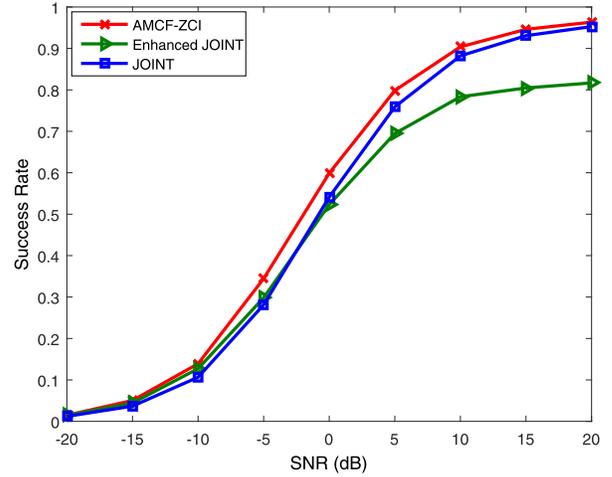
and JOINT, respectively. As shown in Fig. 5, we can see that the performance of AMCF-ZCI is better than that of JOINT and EJOINT. The reason on the performance improvement of AMCF-ZCI over JOINT is that AMCF-ZCI uses all the antennas while half of antennas may be powered off by JOINT. The reason on the performance improvement of AMCF-ZCI over EJOINT is that AMCF-ZCI has better beam pattern than EJOINT. Note that EJOINT performs better than JOINT at lower signal-to-noise-ratio (SNR) region and performs worse than JOINT at high SNR region, because EJOINT has a worse beam pattern although it uses all the antennas.

B. Evaluation of Simultaneous Multiuser Beam Training

We consider a multiuser mmWave massive MIMO system, where the BS equipped with $N_{\text{BS}} = 128$ antennas serves $K = 8$ UEs. Each UE is equipped with $N_{\text{UE}} = 16$ antennas. The mmWave MIMO channel matrix is assumed to have $L_k = 3$ channel paths with one LOS path and two non-line-of-sight (NLOS) paths, where the channel gain of the LOS path obeys $\lambda_1 \sim \mathcal{CN}(0,1)$ and that of the two NLOS paths obeys $\lambda_2 \sim \mathcal{CN}(0, 0.01)$ and $\lambda_3 \sim \mathcal{CN}(0, 0.01)$. Both the physical channel AoA ω_{UE}^l and physical channel AoD ω_{BS}^l of the lth channel path for $l = 1, 2, 3$ obey the uniform distribution over $[0, 2\pi]$ [14], [18]. According to the discussion in the previous subsection, we know that AMCF-ZCI has the best performance among the four schemes taken into comparison. Therefore, we use AMCF-ZCI to design the hierarchical codebook for each UE.

Fig. 6 compares the success rate of beam training for different schemes. Since there are totally K UEs served by the BS, the success rate shown in Fig. 6 is averaged over all K UEs. To make fair comparison, we first extend the scheme in [15] from partially connected structure to fully connected structure. It is seen that the scheme in [11] can achieve better performance than the other three schemes, which lies in the fact that the beam sweeping inherently performs better than the hierarchical beam training. Note that in order to clearly present our idea in this work, we start the hierarchical beam training from the top layer of the hierarchical codebook for both the
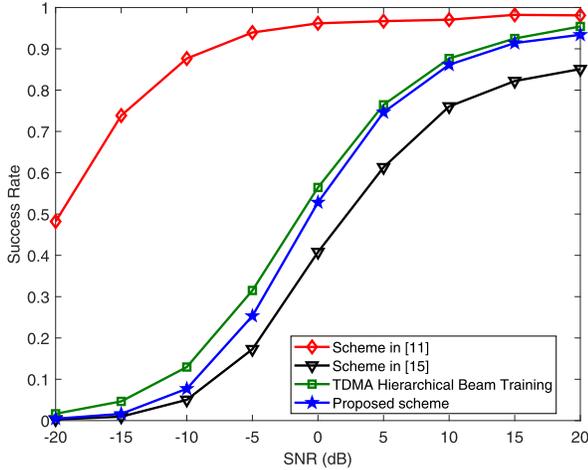
Fig. 6. Comparisons of success rate of beam training for different schemes.
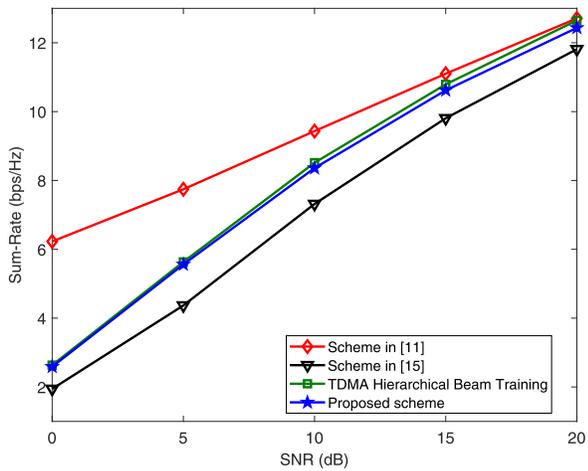
Fig. 7. Comparisons of the averaged sum-rate for different schemes.

BS and the UEs. In fact, we may start the hierarchical beam training from the lower layer of the hierarchical codebook to enlarge the beam gain of the codewords, which can improve the beam training performance. Since the training overhead of the scheme in [11] is much higher than the other schemes, our interest is indeed the comparison of the three hierarchical beam training schemes. It is seen that the performance of our scheme is better than the scheme in [15] and almost the same as that of TDMA hierarchical beam training. At low SNR region, e.g., SNR = −5dB, our scheme performs slightly worse than the TDMA hierarchical beam training, which is caused by the lower signal power averaged over all UEs in the simultaneous beam training of our scheme. However, the training overhead of our scheme is much smaller than the TDMA hierarchical beam training, i.e., 176 versus 36 with 79.6% reduction.

Fig. 7 compares the averaged sum-rate for different beam training schemes. It is seen that the curves of our scheme and the TDMA hierarchical beam training scheme are almost overlapped. Moreover, as the SNR increases, the performance gap between our scheme and the scheme in [11] gets smaller. At SNR = 15dB, the gap is no more than 0.5 bps/Hz, which demonstrates that our scheme can approach the performance of the beam sweeping with the considerable reduction in training overhead.

VI. CONCLUSION

In this article, we have considered multiuser beam training based on hierarchical codebook for mmWave massive MIMO, where the BS can simultaneously perform the beam training with multiple UEs. For the UEs, we have proposed AMCF-ZCI to design the hierarchical codebook with constant modulus constraint. For the BS, we have designed the hierarchical codebook in an adaptive manner, where the codewords in the current layer are designed according to the beam training results of the previous layer. In particular, we have designed multi-mainlobe codewords for the BS, where each mainlobe of the multi-mainlobe codeword covers a spatial region that one or more UEs are probably in. Except for the bottom layer, there are only two codewords at each layer in the designed adaptive hierarchical codebook, which only requires twice simultaneous beam training for all the UEs no matter how many UEs the BS serves. Simulation results have verified the effectiveness of the proposed hierarchical codebook design schemes and have shown that the proposed simultaneous multiuser beam training scheme can approach the performance of the beam sweeping but with considerable reduction in beam training overhead. Our future work will focus on the reduction of feedback from the UEs to the BS during the multiuser beam training as well as the extension of our beam training scheme from the narrowband mmWave channel model to the wideband one.

REFERENCES

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

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

[3] L. Zhao, G. Geraci, T. Yang, D. W. K. Ng, and J. Yuan, "A tone-based AoA estimation and multiuser precoding for millimeter wave massive MIMO," *IEEE Trans. Commun.*, vol. 65, no. 12, pp. 5209–5225, Dec. 2017.

[4] C. Lin, G. Y. Li, and L. Wang, "Subarray-based coordinated beamforming training for mmwave and sub-THz communications," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 9, pp. 2115–2126, Sep. 2017.

[5] Z. Xiao, L. Zhu, J. Choi, X. Chao, and X.-G. Xia, "Joint power allocation and beamforming for non-orthogonal multiple access (NOMA) in 5G millimeter wave communications," *IEEE Trans. Wireless Commun.*, vol. 17, no. 5, pp. 2961–2974, May 2018.

[6] Z. Wei, L. Zhao, J. Guo, D. W. K. Ng, and J. Yuan, "Multi-beam NOMA for hybrid mmwave systems," *IEEE Trans. Commun.*, vol. 67, no. 2, pp. 1705–1719, Feb. 2019.

[7] B. Wang, F. Gao, S. Jin, H. Lin, and G. Y. Li, "Spatial- and frequency-wideband effects in millimeter-wave massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 66, no. 13, pp. 3393–3406, Jul. 2018.

[8] 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.

[9] 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.

[10] A. Ali, N. Gonzalez-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.

[11] 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.

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

[13] 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.

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

[15] 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. IEEE Global Commun. Conf. (GLOBECOM)*, Abu Dhabi, UAE, Dec. 2018, pp. 1–6.

[16] C. Qi, P. Dong, W. Ma, H. Zhang, Z. Zhang, and G. Ye Li, "Acquisition of channel state information for mmwave massive MIMO: Traditional and machine learning-based approaches," 2020, *arXiv:2006.08894*. [Online]. Available: http://arxiv.org/abs/2006.08894

[17] Z. Xiao, H. Dong, L. Bai, P. Xia, and X.-G. Xia, "Enhanced channel estimation and codebook design for millimeter-wave communication," *IEEE Trans. Veh. Technol.*, vol. 67, no. 10, pp. 9393–9405, Oct. 2018.

[18] W. Ma and C. Qi, "Beamspace channel estimation for millimeter wave massive MIMO system with hybrid precoding and combining," *IEEE Trans. Signal Process.*, vol. 66, no. 18, pp. 4839–4853, Sep. 2018.

[19] O. Aldayel, V. Monga, and M. Rangaswamy, "Tractable transmit MIMO beampattern design under a constant modulus constraint," *IEEE Trans. Signal Process.*, vol. 65, no. 10, pp. 2588–2599, May 2017.

[20] W. Fan, C. Zhang, and Y. Huang, "Flat beam design for massive MIMO systems via Riemannian optimization," *IEEE Wireless Commun. Lett.*, vol. 8, no. 1, pp. 301–304, Feb. 2019.

[21] R. Peng and Y. Tian, "Robust wide-beam analog beamforming with inaccurate channel angular information," *IEEE Commun. Lett.*, vol. 22, no. 3, pp. 638–641, Mar. 2018.

[22] D. Chu, "Polyphase codes with good periodic correlation properties (Corresp.)," *IEEE Trans. Inf. Theory*, vol. 18, no. 4, pp. 531–532, Jul. 1972.

[23] S. Noh, M. D. Zoltowski, and D. J. Love, "Multi-resolution codebook and adaptive beamforming sequence design for millimeter wave beam alignment," *IEEE Trans. Wireless Commun.*, vol. 16, no. 9, pp. 5689–5701, Sep. 2017.

[24] 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.

[25] 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, pp. 170–180, 2020.

[26] S. He, J. Wang, Y. Huang, B. Ottersten, and W. Hong, "Codebook-based hybrid precoding for millimeter wave multiuser systems," *IEEE Trans. Signal Process.*, vol. 65, no. 20, pp. 5289–5304, Oct. 2017.

Kangjian Chen (Member, IEEE) received the B.S. degree from the Nanjing University of Science and Technology, Nanjing, China, in 2017, and the M.S. degree in signal processing from Southeast University, Nanjing, China, in 2020. His research interest includes signal processing algorithms for millimeter-wave and massive MIMO communications. In 2019, he received the IEEE Global Communications Conference (GLOBECOM) Best Paper Award.

Octavia A. Dobre (Fellow, IEEE) received the Dipl.Ing. and Ph.D. degrees from the Polytechnic Institute of Bucharest, Romania, in 1991 and 2000, respectively.

From 2002 to 2005, she was with the New Jersey Institute of Technology, Newark, NJ, USA. In 2005, she joined Memorial University, St. John's, NL, Canada, where she is currently a Professor and a Research Chair. She was a Visiting Professor with the Massachusetts Institute of Technology, Cambridge, MA, USA, and the Universit de Bretagne Occidentale, France. Her research interests encompass various wireless technologies, such as nonorthogonal multiple access and full duplex, as well as optical and underwater communications. She has coauthored more than 300 refereed journal articles in these areas.

Dr. Dobre is a fellow of the Engineering Institute of Canada. She obtained the best paper awards at various conferences, including the IEEE ICC, the IEEE GLOBECOM, and the IEEE WCNC. She has also served as the General Chair, the Technical Program Co-Chair, the Tutorial Co-Chair, and the Technical Co-Chair of Symposia at numerous conferences. She was a Royal Society Scholar and a Fulbright Scholar. She was the Editor-in-Chief (EiC) of the IEEE COMMUNICATIONS LETTERS, a Senior Editor, an Editor, and a Guest Editor for various prestigious journals and magazines. She serves as the EiC of the IEEE Open Journal of the Communications Society. She is a Distinguished Lecturer of the IEEE Communications Society.

Chenhao Qi (Senior Member, IEEE) received the B.S. degree (Hons.) in information engineering from the Chien-Shiung Wu Honored College, Southeast University, in 2004, and the Ph.D. degree majored in signal and information processing from Southeast University in 2010.

From 2008 to 2010, he visited the Department of Electrical Engineering, Columbia University, New York, NY, USA. Since 2010, he has been with the faculty of the School of Information Science and Engineering, Southeast University, where he is currently an Associate Professor. His research interests include millimeter-wave communications, massive MIMO, satellite communications, and intelligent signal processing. In 2019, he received the IEEE Global Communications Conference (GLOBECOM) Best Paper Award as well as the Eleventh International Conference on Wireless Communications and Signal Processing (WCSP) Best Paper Award. He is the IEEE COMMUNICATIONS LETTERS Exemplary Reviewer of 2017 and the IEEE COMMUNICATIONS LETTERS Exemplary Editor of 2018. He is also an IEEE ACCESS Outstanding Associate Editor of 2018. He serves as an Associate Editor for the IEEE TRANSACTIONS ON COMMUNICATIONS, the IEEE COMMUNICATIONS LETTERS, IEEE ACCESS, the IEEE OPEN JOURNAL OF THE COMMUNICATIONS SOCIETY, and *China Communications*. He serves as a Symposium Co-Chair for international conferences, including GLOBECOM, WCSP, and the IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC).

Geoffrey Ye Li (Fellow, IEEE) was with AT&T Labs Research, Red Bank, NJ, USA, as a Senior and then a Principal Technical Staff Member, from 1996 to 2000. He was a Post-Doctoral Research Associate with the University of Maryland, College Park, MD, USA, from 1994 to 1996. He is currently a Professor with the Georgia Institute of Technology, Atlanta, GA, USA. His general research interests include statistical signal processing and machine learning for wireless communications. In these areas, he has published more than 500 journal articles and conference papers in addition to more than 40 granted patents. His publications have been cited more than 42 000 times, and he has been recognized as the World's Most Influential Scientific Mind, also known as a Highly Cited Researcher, by Thomson Reuters almost every year.

Dr. Li was awarded IEEE Fellow for his contributions to signal processing for wireless communications in 2005. He won several prestigious awards from the IEEE Signal Processing Society (Donald G. Fink Overview Paper Award in 2017), the IEEE Vehicular Technology Society (James Evans Avant Garde Award in 2013 and the Jack Neubauer Memorial Award in 2014), and the IEEE Communications Society (Stephen O. Rice Prize Paper Award in 2013, the Award for Advances in Communication in 2017, and the Edwin Howard Armstrong Achievement Award in 2019). He also received the 2015 Distinguished Faculty Achievement Award from the School of Electrical and Computer Engineering, Georgia Tech. He has been involved in editorial activities for more than 20 technical journals, including the founding Editor-in-Chief of IEEE 5G TECH FOCUS and the Founding Editor-in-Chief of the IEEE JSAC Special Series on ML in Communications and Networking. He has organized and chaired many international conferences, including Technical Program Vice-Chair of the IEEE ICC'03, the Technical Program Co-Chair of the IEEE SPAWC'11, the General Co-Chair of the IEEE GlobalSIP'14, the Technical Program Co-Chair of the IEEE VTC'16 (Spring), and the General Co-Chair of the IEEE VTC'19 (Fall) and the IEEE SPAWC'20.