Simultaneous Multi-Beam Training for Millimeter-Wave Communication System

Yinxiao Zhuo, Ziyuan Sha, Zhaocheng Wang, *Fellow, IEEE*,
and Sheng Chen, *Fellow, IEEE*

***Abstract*—Millimeter-wave communication is regarded as a promising technology for the next generation communication system, where beamforming and beam training are essential for power concentration and coverage extension. In this paper, we propose an angular domain separation based multi-beam training (ADS-MBT) scheme and an angular domain separation based hierarchical multi-beam training (ADS-HMBT) scheme, which enable simultaneous multi-beam training under the single data stream constraint. We provide theoretical analysis on the performance of the proposed schemes and optimize the non-orthogonal codebook used in the proposed schemes accordingly. With this appropriate design, our proposed ADS-MBT scheme can achieve almost the same beamforming gain as the exhaustive scheme but only imposing half or less time slot consumption. The proposed ADS-HMBT scheme further reduces the training overhead, and it outperforms the existing two-layer hierarchical search, in terms of training overhead and beamforming gain in the multiple feedback beam training case. Simulation results show the effectiveness and superior performance of our both proposed schemes over the existing state-of-the-arts.**

***Index Terms*—Millimeter-wave, beam training, initial access, angular domain separation.**

I. Introduction

Due to the shortage of available sub-6GHz spectrum, millimeter-wave (mmWave) band has attracted considerable attention. With abundant spectrum resources, mmWave has the potential to support communication of gigabits-per-second data rates [1]–[4]. However, mmWave wireless communication suffers high path loss and molecular absorption [5], [6]. Therefore, large-scale antenna arrays are required for power concentration and capacity improvement [7].

To optimize the beamforming gains of large-scale array, base station (BS) and user equipment (UE) need to perform beam training procedure, in which BS and UE search for the best pair of transmitting and receiving beamforming vectors, respectively. In practical implementation, codebook-based beam training is preferred, where UE measures several reference signals corresponding to several beams in a predefined codebook and feeds back the index of the selected beam. Generally, codebook size is proportional to antenna number to ensure spatial coverage. Therefore, for large-scale array in mmWave communication, exhaustive search as illustrated in Fig. 1 (a), although guaranteeing to select the optimal beam, introduces unacceptable beam training overhead.

To reduce beam training overhead, many improved beam training schemes have been proposed. One common solution is to perform hierarchical search with multi-resolution beamforming codebook [8]–[13]. Specifically, as illustrated in Fig. 1 (b), BS exhaustively activates wide beams to identify a coarse direction of UE, and then gradually narrows down the range of UE direction according to UE's feedback until UE selects the optimal narrow beam. Under these hierarchical search schemes, however, the beam refinement process cannot be shared by the UE of different wide beam ranges. Generally, BS needs to perform beam refinement for each UE individually, which introduces extra overhead. Moreover, increasing the number of layers increases feedback overhead, which introduces large delay and makes this scheme inapplicable for initial access stage (see Appendix). Hence, only two-layer hierarchical search is accepted in practice, but this degrades the performance in terms of training overhead and beamforming gains. Another popular solution is to exploit the measurements on wide beams to estimate the optimal narrow beam [14], [15]. However, these works are prone to errors in practical noisy multipath channel due to under-exploration of the beam space. Additionally, machine learning assisted beam training

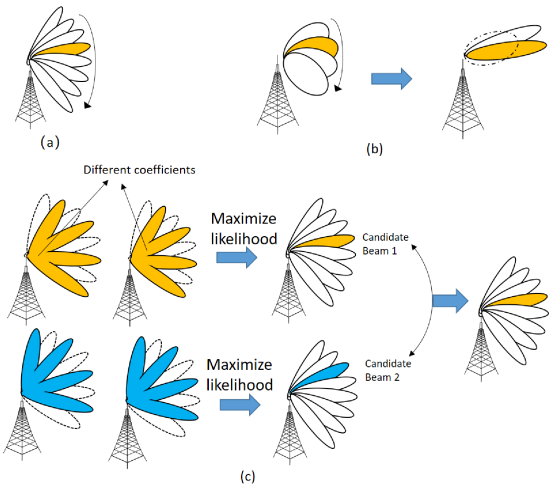


Fig. 1. Illustrations of different beam training schemes. (a) Exhaustive search: BS sequentially activates all the beams one by one, and UE selects the optimal beam. (b) Hierarchical search: BS first activates wide beams and then performs narrow beam refinement. (c) Proposed ADS-MBT scheme: BS divide beams into several groups and transmits merged beams with different coefficients for several times. UE firstly selects the candidate optimal beam in each group, and then compare all the candidate beams to select the optimal one.

This work was supported in part by the National Key R&D Program of China under Grant 2018YFB1801102 and in part by National Natural Science Foundation of China (Grant No. 61871253). *(Corresponding author: Zhaocheng Wang.)*

Y. Zhuo, Z. Sha and Z. Wang are with the Beijing National Research Center for Information Science and Technology, Department of Electronic Engineering, Tsinghua University, Beijing 100084, China (E-mails: zhuoyx20@mails.tsinghua.edu.cn, shazy17@mails.tsinghua.edu.cn, zcwang@tsinghua.edu.cn). Z. Wang is also with the Division of Information Science and Technology, Graduate School at Shenzhen, Tsinghua University, Shenzhen 518055, China.

S. Chen is with the School of Electronics and Computer Science, University of Southampton, Southampton SO17 1BJ, U.K. (E-mail: sqc@ecs.soton.ac.uk).

also offers a potential solution for the optimal beam decision [16]–[18]. However, a machine learning trained model is only valid to the specific channel that it was trained for, and it cannot be easily applied or transported to other environments.

In addition to the aforementioned sequential single beam search schemes [8]–[19], some researchers investigated the possibility of simultaneously training multiple beams [20]–[22]. Specifically, in [20], both BS and UE are equipped with multiple radio frequency (RF) chains to support multi-stream transmission, and hence multiple beams can be simultaneously measured via parallel multi-stream transmissions. The works [21], [22] proposed bisection search schemes to conduct parallel beam training among UE group. However, on one hand, these schemes can support simultaneous multiple beams training only when there are multiple data streams (links) between BS and UE. In a more practical case where UE has only one RF chain, UE can only receive single data stream. Under this condition, these beam training schemes still sweep one beam in each time slot, and beam training overhead is not reduced. On the other hand, these bisection search schemes perform well only in single-path scenario, and suffer serious performance degradation in multipath environment. Some other schemes use Bayes posterior probability updating to simultaneously evaluate all the beams and select the optimal one [23], [24]. But the Bayes updating process is intensively feedback-based, which also introduces large feedback delay.

It can be seen that the existing beam training schemes have some shortcomings and they are not suitable for practical initial access. Specifically, hierarchical search [8]–[13] and Bayesian philosophy [23], [24] based beam training schemes introduce large feedback delay, since multiple rounds of UE feedback are required. Meanwhile, the machine learning assisted beam training schemes [16]–[18] are sensitive to environment and are computational expensive. To perform robust beam training in practical initial access while reducing the overhead, in this paper, we propose an angular domain separation (ADS) based multi-beam training (ADS-MBT) scheme, to support simultaneously multi-beam training under the single data stream constraint. Our basic idea is illustrated in Fig. 1 (c). Specifically, the proposed ADS-MBT scheme divides beams into several angular-domain separated beam groups. Due to angular concentration of mmWave channel, at most one beam in each group has high beamforming gain. In the merged beam sweeping stage, beams in each group are activated simultaneously with different coefficient combinations for several times. At UE side, UE performs maximum likelihood (ML) detection to recognize the candidate optimal beam in each group and estimates its beamforming gain. Then UE compares all the candidate beams to select the optimal beam. Moreover, we extend the ADS-MBT to a multiple feedback scheme, called ADS-hierarchical multi-beam training (ADS-HMBT), to further reduce training overhead at the cost of introducing extra feedback. The proposed ADS-MBT and ADS-HMBT schemes can achieve desired and flexible tradeoffs between beam training performance and overhead by simply adjusting merged beam sweeping times. Additionally, we provide theoretical analysis on the performance of the proposed schemes and accordingly optimize the non-orthogonal merging coefficient matrix (MCM) design. Numerical simulation results show that our schemes outperform existing beam training counterparts in terms of training overhead and/or beamforming gains. Our main contributions are summarized as follows.

- Two multi-beam training schemes, ADS-MBT and ADS-HMBT, are proposed. The both schemes support simultaneously multi-beam training under the single data stream constraint, thus significantly reducing the training overhead. Our schemes outperform the exhaustive search and hierarchical search schemes as well as other existing state-of-the-art schemes.
- In theoretical analysis, we derive beam misalignment probability approximation, and accordingly provide two alternative MCM designs, which ensures that our proposed schemes perform well under different channel conditions.
- In addition to reduce training overhead significantly, our proposed schemes can flexibly set the desired tradeoff between training performance and overhead according to different application scenarios, which enhances the quality of service and spectral efficiency.

The rest of the paper is organized as follows. In Section II, the geometrical mmWave channel model is given and the codebook based beam training is formulated as an optimization problem. Section III details our ADS-MBT beam training scheme and its extension the ADS-HMBT scheme. Theoretic performance analysis is provided in Section IV, while the MCM design is derived in Section V. Section VI offers numerical simulation results to demonstrate the superior performance of our proposed schemes. Our conclusions are drawn in Section VII.

Throughout our discussion, the following notation conventions are adopted. Boldface capital and lower-case letters stand for matrices and vectors, respectively, e.g., $\boldsymbol{A}$ and $\boldsymbol{a}$. The transpose, Hermitian transpose and inverse operators are denoted by $(\cdot)^{\mathrm{T}}$, $(\cdot)^{\mathrm{H}}$ and $(\cdot)^{-1}$, respectively, while $\boldsymbol{A}(i,j)$, $\boldsymbol{A}(i,:)$ and $\boldsymbol{A}(:,j)$ denote the (i,j)-th element of $\boldsymbol{A}$, the i-th row vector of $\boldsymbol{A}$ and the j-th column vector of $\boldsymbol{A}$, respectively. Two-norm of $\boldsymbol{a}$ is denoted by $\|\boldsymbol{a}\|_2$, and $|a|$ denotes amplitude of a. Cursive capital letters denote sets, e.g., $\mathcal{A}$, and $|\mathcal{A}|$ denotes the cardinality of set $\mathcal{A}$, while $\boldsymbol{e}_n$ denotes the unit vector whose n-th element is 1 and all the other elements are zero. The $M \times M$ identity matrix is denoted as $\boldsymbol{I}_M$, and $\otimes$ is the Kronecker-product operator.

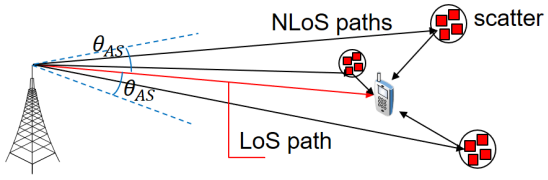


Fig. 2. Typical mmWave multi-path channel with one LoS path and multiple NLoS paths, where θ_{AS} denotes the angular spread.

II. System Model

A. MmWave Multipath Channel

We consider a typical multipath channel with single UE receiver, where BS and UE are equipped with N_T and N_R antennas, respectively. Furthermore, uniform linear array (ULA) with half-wavelength element spacing are utilized. In consideration of power consumption, hybrid architecture with limited RF chains is preferred, and Subsection V-A shows that even single-RF-chain fully analog architecture is applicable with appropriate design. The channel can be assumed invariant since the beam training duration is short. The downlink channel matrix $\boldsymbol{H} \in \mathbb{C}^{N_R \times N_T}$ is the composition of one line-of-sight (LoS) path and P non-line-of-sight (NLoS) paths as shown in Fig. 2, which is expressed as [25], [26]

$$\boldsymbol{H} = \sum_{p=0}^{P} \gamma_p \boldsymbol{H}_p, \tag{1}$$

where $\boldsymbol{H}_0 \in \mathbb{C}^{N_R \times N_T}$ represents the LoS path, and $\boldsymbol{H}_p \in \mathbb{C}^{N_R \times N_T}$, $1 \le p \le P$, are the P NLoS paths, while $\gamma_p \in \mathbb{C}$, $p = 0, 1, \cdots, P$, are the complex gains for these paths. In a typical LoS dominated scenario, the LoS path will be much stronger than all the NLoS paths, which suggests that $|\gamma_0| \gg |\gamma_p|$, $1 \le p \le P$, while in a typical multipath scenario, $|\gamma_0|$ and $|\gamma_p|$, $1 \le p \le P$, are comparable. Denoting $\theta_p^{(AoD)}$ and $\theta_p^{(AoA)}$ as the angle of departure (AoD) and the angle of arrival (AoA) of the p-th path, respectively, then $\boldsymbol{H}_p$ can be expressed as

$$\boldsymbol{H}_p = \sqrt{N_T N_R}\, \boldsymbol{a}_R\big(\theta_p^{(AoA)}\big) \boldsymbol{a}_T^{\mathrm{H}}\big(\theta_p^{(AoD)}\big), \tag{2}$$

where $\boldsymbol{a}_T\big(\theta^{(AoD)}\big) \in \mathbb{C}^{N_T}$ and $\boldsymbol{a}_R\big(\theta^{(AoA)}\big) \in \mathbb{C}^{N_R}$ denote the transmitting and receiving antenna response vectors with azimuth angles $\theta^{(AoD)}$ and $\theta^{(AoA)}$, respectively, which are given by

$$\boldsymbol{a}_T(\theta) = \frac{1}{\sqrt{N_T}} \big[1 \; e^{-\mathrm{j}\pi\cos\theta} \cdots e^{-(N_T-1)\mathrm{j}\pi\cos\theta}\big]^{\mathrm{T}}, \tag{3}$$

$$\boldsymbol{a}_R(\theta) = \frac{1}{\sqrt{N_R}} \big[1 \; e^{-\mathrm{j}\pi\cos\theta} \cdots e^{-(N_R-1)\mathrm{j}\pi\cos\theta}\big]^{\mathrm{T}}. \tag{4}$$

Generally, since BS commonly locates at high position, the multipath effect mainly arises due to obstacles around UE. Therefore, the multi-path components exhibit angular concentration, i.e., the AoDs of the NLoS paths are closed to the AoD of the LoS path. Denoting θ_{AS} as the angular spread, which is the root mean square (RMS) of angle difference between LoS and NLoS AoDs, we have

$$\theta_{AS} = \sqrt{\frac{1}{P}\sum_{p=1}^{P}\Big(\theta_p^{(AoD)} - \theta_0^{(AoD)}\Big)^2}. \tag{5}$$

Generally, θ_{AS} is smaller than 10° for mmWave channels [27], [28]. Due to the angular concentration property, the BS beam with direction far away from $\theta_0^{(AoD)}$ has low beamforming gain.

With the above mmWave channel model, single data stream transmission is formulated as

$$y = \sqrt{\beta}\boldsymbol{w}^{\mathrm{H}}\boldsymbol{H}\boldsymbol{f}s + \boldsymbol{w}^{\mathrm{H}}\boldsymbol{n}, \tag{6}$$

where $s \in \mathbb{C}$ is the transmitted symbol with β denoting the transmission power, $\boldsymbol{f} \in \mathbb{C}^{N_T \times 1}$ with $\|\boldsymbol{f}\|_2 = 1$ represents the transmitter beamforming vector, and $\boldsymbol{n} \in \mathbb{C}^{N_R \times 1}$ is the additive Gaussian white noise (AWGN) vector at UE, while $\boldsymbol{w} \in \mathbb{C}^{N_R \times 1}$ with $\|\boldsymbol{w}\|_2 = 1$ represents the receiver beamforming or combining vector, and y is the output of the UE receiving beamformer, which is the sufficient statistic for detecting s.

B. Codebook Based Beam Training

Denoting the effective channel $h_{eff} = \boldsymbol{w}^{\mathrm{H}}\boldsymbol{H}\boldsymbol{f}$, (6) can be rewritten as

$$y = \sqrt{\beta}h_{eff}s + \widetilde{n}, \tag{7}$$

where $\widetilde{n} = \boldsymbol{w}^{\mathrm{H}}\boldsymbol{n}$ is the equivalent noise and $\big|h_{eff}\big|^2$ can be regarded as the beamforming gain. In beam training procedure, BS and UE jointly select the appropriate beamforming vectors $\boldsymbol{f}$ and $\boldsymbol{w}$ to achieve the optimal beamforming gain, which can be formulated as

$$\big\{\boldsymbol{f}^\star, \boldsymbol{w}^\star\big\} = \arg \max_{\|\boldsymbol{f}\|_2=1, \|\boldsymbol{w}\|_2=1} \big|\boldsymbol{w}^{\mathrm{H}}\boldsymbol{H}\boldsymbol{f}\big|. \tag{8}$$

Based on singular value decomposition (SVD), the best transmitting and receiving beamforming vectors are given by the principle right and left singular vectors of the channel matrix $\boldsymbol{H}$ [29], respectively. Specifically, let the SVD of $\boldsymbol{H}$ be $\boldsymbol{H} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{H}}$ with the singular values arranged in decreasing order. Then the best beamforming vectors are $\boldsymbol{f}^\star = \boldsymbol{V}(:,1)$ and $\big(\boldsymbol{w}^\star\big)^{\mathrm{T}} = \boldsymbol{U}(:,1)$.

However, the acquisition of the full channel state information, i.e., the channel matrix $\boldsymbol{H}$, requires high pilot overhead and feedback overhead. Hence codebook based beam training is introduced to facilitate practical application of beamforming before channel estimation. Specifically, $\boldsymbol{f}$ and $\boldsymbol{w}$ are selected from the predefined codebook sets $\mathcal{F}$ and $\mathcal{W}$,

$$\big\{\boldsymbol{f}^\star, \boldsymbol{w}^\star\big\} = \arg \max_{\boldsymbol{f}\in\mathcal{F}, \boldsymbol{w}\in\mathcal{W}} \big|\boldsymbol{w}^{\mathrm{H}}\boldsymbol{H}\boldsymbol{f}\big|. \tag{9}$$

We consider the most widely used DFT codebook in our discussions. For other beam codebook, our proposed schemes are still applicable as long as the beam grouping satisfies ADS property. In the standard DFT codebook without oversampling, the number of orthogonal beams equal to the number of antennas which cover the entire angular space. the DFT codebook can be expressed as

$$\mathcal{F} = \left\{\boldsymbol{a}_T\big(\theta_k^T\big) \,\middle|\, \cos\theta_k^T = \frac{2k-1-N_T}{N_T}, k = 1, 2, \cdots, N_T\right\}, \tag{10}$$

$$\mathcal{W} = \left\{\boldsymbol{a}_R\big(\theta_k^R\big) \,\middle|\, \cos\theta_k^R = \frac{2k-1-N_R}{N_R}, k = 1, 2, \cdots, N_R\right\}. \tag{11}$$

III. Proposed Schemes

This section derives our proposed ADS-MBT scheme and the multiple feedback aided ADS-HMBT scheme. We start with a discussion on UE receiving beam.

A. UE Receiving Beam

In general case, UE is equipped with antenna array and it also needs to perform beam selection. But it is sufficient to consider only single-antenna UE. IEEE 802.11ad [30] specifications utilize two schemes, UE exhaustive search and interactive search [31], to separate UE beamforming and BS beamforming. In the UE exhaustive search scheme, the beam training at BS side repeats for $|\mathcal{W}|$ times, and in each time UE uses a different beam to receive. In the interactive search scheme, UE first activates only one antenna to perform BS side beam training, and then BS fixes the optimal beam and perform UE side beam training, which can be formulated as

$$\boldsymbol{f}^{\star} = \arg\max_{\boldsymbol{f}\in\mathcal{F}} |\boldsymbol{H}(1,:)\boldsymbol{f}|, \tag{12}$$

$$\boldsymbol{w}^{\star} = \arg\max_{\boldsymbol{w}\in\mathcal{W}} |\boldsymbol{w}^{\mathrm{H}}\boldsymbol{H}\boldsymbol{f}^{\star}|. \tag{13}$$

Since the both schemes in the specifications [30] imply that UE side beam can be assumed fixed when focusing on BS side beam training, our beam training optimization problem can be deduced to be

$$\boldsymbol{f}^{\star} = \arg\max_{\boldsymbol{f}\in\mathcal{F}} \left|\boldsymbol{h}_{eff}^{\mathrm{H}}\boldsymbol{f}\right|, \tag{14}$$

where $\boldsymbol{h}_{eff} = \boldsymbol{H}^{\mathrm{H}}\boldsymbol{w}$ is the equivalent channel with the fixed UE beam $\boldsymbol{w}$. We note that (14) is equivalent to the single-antenna ($N_R = 1$) UE case. Therefore, without loss of generality, single-antenna UE is considered in our derivations and discussions. The degenerated channel vector with $N_R = 1$ $\boldsymbol{h} \in \mathbb{C}^{N_T\times 1}$ is given by

$$\boldsymbol{h} = \sum_{p=0}^{P} \gamma_p \sqrt{N_T}\boldsymbol{a}_T\left(\theta_p^{(AoD)}\right). \tag{15}$$

B. ADS-MBT

The proposed ADS-MBT scheme consists of the three steps: beam merging, repeated transmission and beam recognition. In the beam merging stage, BS divides all the N_T beams into several groups to satisfy the ADS property. In the repeated transmission stage, BS sends each beam group with different beam merging coefficients for several times. Finally, in the beam recognition stage, UE uses the ML detection to recognizes the candidate optimal beam in each group, and then compares all the candidate beams to select the optimal narrow beam.

1) Beam Merging: BS divides all the N_T beams into several groups and merges the beams in each group into a merged beam. To utilize the AoD concentration of mmWave multipath channel and facilitate the subsequent beam recognition, the beams in each group should be far away from each other.

Let the N_T beams be divided into G groups, and the g-th group consists of N_g beams, $1 \le g \le G$. Denote the beams in the g-th group by $\mathcal{F}_g = \left\{\boldsymbol{f}_{g,1}, \boldsymbol{f}_{g,2}, \cdots, \boldsymbol{f}_{g,N_g}\right\} \subset \mathcal{F}$, where the beamforming vectors $\boldsymbol{f}_{g,n_g} = \boldsymbol{a}_T\left(\theta_{g,n_g}^T\right)$, $n_g = 1, 2, \cdots, N_g$, and further define the g-th merged beam matrix as $\boldsymbol{F}_g = \left[\boldsymbol{f}_{g,1}\ \boldsymbol{f}_{g,2}\cdots\boldsymbol{f}_{g,N_g}\right] \in \mathbb{C}^{N_T\times N_g}$, $1 \le g \le G$. Obviously, we have

$$\mathcal{F}_i \bigcap \mathcal{F}_j = \emptyset,\ 1 \le i < j \le G, \tag{16}$$

$$\bigcup_{g=1}^{G} \mathcal{F}_g = \mathcal{F}. \tag{17}$$

A reasonable grouping method is to choose

$$\mathcal{F}_g = \left\{\boldsymbol{a}_T\left(\theta_k^T\right) \middle| k \equiv g \pmod{G}\right\}. \tag{18}$$

Note that we have

$$\left|\theta_{g,i}^T - \theta_{g,j}^T\right| \ge 2G \arcsin\left(\frac{1}{N_T}\right) \ge \frac{2G}{N_T}. \tag{19}$$

When $G \ge \theta_{AS}N_T$, the grouping method (18) satisfies the following ADS property

$$\left|\theta_{g,i}^T - \theta_{g,j}^T\right| \ge 2\theta_{AS}, 1 \le i < j \le N_g, 1 \le g \le G. \tag{20}$$

With this ADS property and the angular concentration property of AoDs, at most one beam in each group has high beamforming gain, while the beamforming gains of all the other beams are much lower and negligible. Therefore, UE can exploit this property to distinguish the optimal beam from simultaneously transmitted beams.

Denote the optimal beam in the g-th group by $\boldsymbol{f}_{g,n_g^\star}$, i.e.,

$$n_g^\star = \arg\max_{1\le n_g\le N_g} \left|\boldsymbol{h}^{\mathrm{H}}\boldsymbol{f}_{g,n_g}\right|, 1 \le g \le G. \tag{21}$$

If $\boldsymbol{f}_{g,n_g^\star}$ has high beamforming gain, we have the following approximation according to the ADS property

$$\boldsymbol{h}^{\mathrm{H}}\boldsymbol{F}_g \approx q_{g,n_g^\star}\boldsymbol{e}_{n_g^\star}^{\mathrm{T}}, 1 \le g \le G, \tag{22}$$

where $q_{g,n_g^\star} = \boldsymbol{h}^{\mathrm{H}}\boldsymbol{f}_{g,n_g^\star}$. If all the beams in the g-th group have low gains, then $q_{g,n_g^\star} \approx 0$, (22) is still a suitable approximation.

After the beam merging stage, BS obtains the G merged beam matrices $\{\boldsymbol{F}_g\}_{g=1}^G$. The merged beam for $\mathcal{F}_g$ realization can be formulated as

$$\bar{\boldsymbol{f}}_g = \sum_{i=1}^{N_g} \boldsymbol{f}_{g,i}s_i = \boldsymbol{F}_g\boldsymbol{s}_g, \tag{23}$$

where $\boldsymbol{s}_g \in \mathbb{C}^{N_g\times 1}$ is the beam merging coefficient vector with $\|\boldsymbol{s}_g\|_2 = 1$, and $\bar{\boldsymbol{f}}_g \in \mathbb{C}^{N_T\times 1}$ is the merged beamforming vector. Fig. 3 shows an example for $N_T = 8$ and $G = 2$. All the 8 beams are divided into the yellow group and the blue group, and in each slot, the beams in one group will be activated simultaneously to form a wide beam.

2) Repeated Transmission: By single measurement of a merged beam, the beams in a group are indistinguishable to UE. To recognize the optimal narrow beam in a group, BS needs to send the same merged beam set with different beam merging coefficients for several times. Specifically, let the g-th merged beam group $\mathcal{F}_g$ be transmitted for M_g times, with the beam merging coefficients $\boldsymbol{s}_{g,1}, \boldsymbol{s}_{g,2}, \cdots, \boldsymbol{s}_{g,M_g}$, respectively. Then the received signals at UE are given by

$$y_{g,k} = \sqrt{\beta}\boldsymbol{h}^{\mathrm{H}}\boldsymbol{F}_g\boldsymbol{s}_{g,k} + n_k = \sqrt{\beta}\boldsymbol{q}_g^{\mathrm{T}}\boldsymbol{s}_{g,k} + n_k, 1 \le k \le M_g, \tag{24}$$

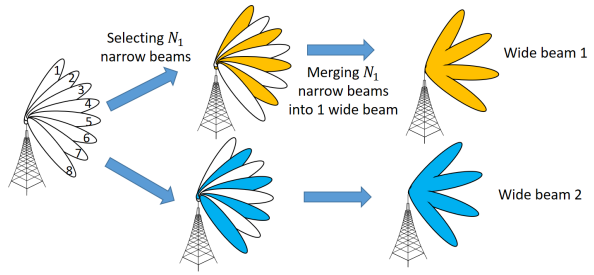


Fig. 3. Beam merging example with $N_T = 8$ and $G = 2$. BS merges 8 narrow beams into 2 wide (merged) beams and activates merged beam in each slot.

where n_k denotes the link's AWGN with power σ_n^2 in the k-th transmission,

$$\boldsymbol{q}_g = \left[q_{g,1}\ q_{g,2}\cdots q_{g,N_g}\right]^{\mathrm{T}} = \left(\boldsymbol{h}^{\mathrm{H}}\boldsymbol{F}_g\right)^{\mathrm{T}} \in \mathbb{C}^{N_g\times 1} \tag{25}$$

is a vector representing the beamforming gains of the beams in $\mathcal{F}_g$ with $q_{g,n_g} = \boldsymbol{h}^{\mathrm{H}}\boldsymbol{f}_{g,n_g}$, $1 \le n_g \le N_g$. The corresponding vector form of the M_g received signals $\boldsymbol{y}_g = \left[y_{g,1}\ y_{g,2}\cdots y_{g,M_g}\right]^{\mathrm{T}}$ can be expressed as

$$\boldsymbol{y}_g = \sqrt{\beta}\left(\boldsymbol{q}_g^{\mathrm{T}}\boldsymbol{S}_g\right)^{\mathrm{T}} + \boldsymbol{n}, 1 \le g \le G, \tag{26}$$

where $\boldsymbol{S}_g = \left[\boldsymbol{s}_{g,1}\ \boldsymbol{s}_{g,2}\cdots\boldsymbol{s}_{g,M_g}\right] \in \mathbb{C}^{N_g\times M_g}$ is the beam merging coefficient matrix (MCM) containing the M_g beam merging coefficient vectors, and $\boldsymbol{n}$ is the corresponding AWGN vector.

According to the ADS principle given in (22), the receiving signals can be approximated as

$$\boldsymbol{y}_g \approx \sqrt{\beta}q_{g,n_g^\star}\left(\boldsymbol{S}_g\left(n_g^\star,:\right)\right)^{\mathrm{T}} + \boldsymbol{n}, \tag{27}$$

where $\boldsymbol{S}(n_g,:)$ can be regarded as a merging coefficient vector corresponding to beam $\boldsymbol{f}_{g,n_g}$. Without prior information, we assume $\left\|\boldsymbol{S}(n_g,:)\right\|_2 = \sqrt{\frac{M_g}{N_g}}$, $1 \le n_g \le N_g$. According to (27), UE is capable of recognizing the best narrow beam from the receiving merged beam signals via a well-designed $\boldsymbol{S}_g$.

An example of repeated transmission is illustrated in Fig. 4, where we have $G = 2$, $M_1 = M_2 = 2$, and $\boldsymbol{S}_1 = \boldsymbol{S}_2 = \boldsymbol{S} \in \mathbb{C}^{4\times 2}$. It can be seen that the number of time slots required for this repeated transmission is $\sum_{g=1}^{G} M_g = 4$.

3) Beam Recognition: In the beam recognition stage, based on the collected received signals $\{\boldsymbol{y}_1, \boldsymbol{y}_2, \cdots, \boldsymbol{y}_G\}$, UE needs to recognize the beam set index and intra-group beam index of the optimal beam with the highest beamforming gain, which can be formulated as

$$\left\{g^\star, n_{g^\star}^\star\right\} = \arg\max_{1\le n_g\le N_g, 1\le g\le G}\left|q_{g,n_g}\right|, \tag{28}$$

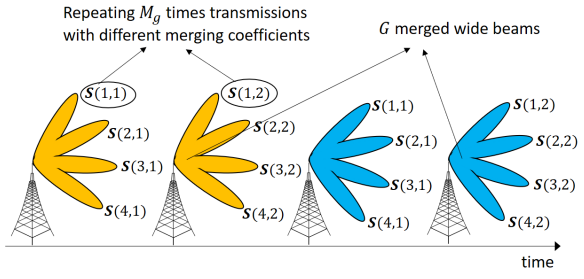


Fig. 4. Repeated transmission example with $M_1 = M_2 = 2$ and $\boldsymbol{S}_1 = \boldsymbol{S}_2 = \boldsymbol{S}$. BS activates each merged beam for 2 times with different merging coefficients.

where $g^\star$ is the best merged beam group index, and $n_{g^\star}^\star$ is the optimal intra-group beam index. In other words, UE needs to select the optimal narrow beam $\boldsymbol{f}_{g^\star,n_{g^\star}^\star}$. We derive the optimal intra-group recognition strategy from the ML principle and the optimal beam set recognition strategy from the maximum estimated gain principle. In the intra-group recognition step, UE selects the optimal beam in each group as a candidate beam and estimates its beamforming gain. In the beam set recognition step, UE compares the estimated gains of all the candidate beams and select the optimal beam.

3.1) Intra-group recognition step: For group $\mathcal{F}_g$, (27) implies that the beamforming gain of $\boldsymbol{f}_{g,n_g^\star}$ determines the received signals. The index and beamforming gain of the optimal beam are given as

$$\left\{n_g^\star, \widehat{q}_g^\star\right\} = \arg\min_{q\in\mathbb{C}, 1\le n_g\le N_g}\left\|\boldsymbol{y}_g - q\left(\boldsymbol{S}_g(n_g,:)\right)^{\mathrm{T}}\right\|_2. \tag{29}$$

For a fixed n_g, the optimal q_g can be derived by

$$\widehat{q}_g^\star = \frac{N_g}{M_g}\boldsymbol{y}_g^{\mathrm{T}}\left(\boldsymbol{S}_g(n_g,:)\right)^{\mathrm{H}}. \tag{30}$$

By denoting $\widehat{q}_{g,n_g} = \frac{N_g}{M_g}\boldsymbol{y}_g^{\mathrm{T}}\left(\boldsymbol{S}_g(n_g,:)\right)^{\mathrm{H}}$, the ML evaluation (29) is equivalent to

$$n_g^\star = \arg\max_{1\le n_g\le N_g}\left|\boldsymbol{y}_g^{\mathrm{T}}\left(\boldsymbol{S}_g(n_g,:)\right)^{\mathrm{H}}\right|^2 = \arg\max_{1\le n_g\le N_g}\left|\widehat{q}_{g,n_g}\right|, \tag{31}$$

$$\widehat{q}_g^\star = \frac{N_g}{M_g}\boldsymbol{y}_g^{\mathrm{T}}\left(\boldsymbol{S}_g(n_g^\star,:)\right)^{\mathrm{H}} = \widehat{q}_{g,n_g^\star}. \tag{32}$$

$\boldsymbol{f}_{g,n_g^\star}$ is the candidate beam from the g-th group and $\widehat{q}_g^\star$ is the estimated beamforming gain, where $1 \le g \le G$.

3.2) Beam set recognition step: UE selects the best beam $\boldsymbol{f}_{g^\star,n_{g^\star}^\star}$ according to

$$g^\star = \arg\max_{1\le g\le G}\left|\widehat{q}_g^\star\right| = \arg\max_{1\le g\le G}\left|\frac{N_g}{M_g}\boldsymbol{y}_g^{\mathrm{T}}\left(\boldsymbol{S}_g(n_g^\star,:)\right)^{\mathrm{H}}\right|. \tag{33}$$

In the proposed recognition strategy, $\boldsymbol{S}_g$ needs to satisfy the single transmission constraint $\|\boldsymbol{S}_g(:,m)\|_2 = 1$, $1 \le m \le M_g$, and the beam power fairness constraint $\|\boldsymbol{S}_g(n,:)\|_2 = \sqrt{\frac{M_g}{N_g}}$, $1 \le n \le N_g$. This recognition strategy computes the correlation coefficients $\widehat{q}_{g,n_g}$ of all the N_T beams and selects the optimal beam with the highest correlation.

Algorithm 1 Proposed ADS-MBT scheme

Input: G, $\{\mathcal{F}_g\}_{g=1}^G$, $\{N_g\}_{g=1}^G$, $\{M_g\}_{g=1}^G$, $\{\boldsymbol{S}_g\}_{g=1}^G$;
Output: The optimal beam index $\left\{g^\star, n_{g^\star}^\star\right\}$;

for $g = 1$ to G **do**
  for $k = 1$ to M_g **do**
    BS transmits merged beam $\bar{\boldsymbol{f}}_{g,k} = \boldsymbol{F}_g\boldsymbol{S}_g(:,k)$;
    UE receives signal $y_{g,k}$;
  end for
end for
UE computes (31), (32) and selects the optimal beam for each group;
UE selects the optimal group $g^\star$ according to (33);
UE feeds back $g^\star, n_{g^\star}^\star$ to BS;
BS uses $\boldsymbol{f}_{g^\star,n_{g^\star}^\star}$ for downlink transmission;

4) Summary of ADS-MBT scheme: Our proposed ADS-MBT scheme is summarized in **Algorithm 1**, where it can be seen that $\sum_{g=1}^{G} M_g$ transmissions and only one feedback are needed. Since calculating each $\widehat{q}_{g,n_g}$ requires N_g complex multiplications, the computational complexity of the ADS-MBT scheme is on the order of $\mathsf{O}\left(\sum_{g=1}^{G} N_g M_g\right)$.

C. Multiple Feedback Aided ADS-HMBT Scheme

In this subsection, we extend our ADS-MBT scheme to a multiple feedback aided scheme, called the ADS-HMBT, to further reduce beam training overhead at a cost of extra feedback delay. The beam merging in the ADS-HMBT remains the same as in the ADS-MBT, and only BS repeated transmission and UE recognition are modified so that they become interactive, linked by one feedback. More specifically, BS first activates only part of the merged beams, and UE first recognizes the optimal beam set based on the received signals. After UE feeds back the optimal beam set index, BS only activates the remaining merged beams in the optimal group, and UE then recognizes the optimal narrow beam based on the previous and these further received signals.

1) BS repeated transmission: For the ADS-MBT scheme, signals for a non-optimal group ($g \neq g^\star$) is only used for the optimal beam set selection. Therefore, with the aid of multiple feedback, time slot consumption may be reduced if the optimal beam set can be selected with fewer measurements.

First, for $1 \leq g \leq G$, BS repeatedly transmits the merged beams of group $\mathcal{F}_g$ for K_g times with the MCM $\boldsymbol{S}_g^{(1)} = \left[\boldsymbol{s}_{g,1}\ \boldsymbol{s}_{g,2} \cdots \boldsymbol{s}_{g,K_g}\right] \in \mathbb{C}^{N_g \times K_g}$, where $K_g < M_g$. This is for UE to recognize the optimal beam set. After UE feeds back the optimal group index $g^\star$, BS then only transmits the remaining $\left(M_{g^\star} - K_{g^\star}\right)$ merged beams of group $\mathcal{F}_{g^\star}$, that is, BS ignores the merged beams in all the other non-optimal groups and only repeatedly transmits the merged beams of the optimal group $\mathcal{F}_{g^\star}$ for $\left(M_{g^\star} - K_{g^\star}\right)$ times with the MCM $\boldsymbol{S}_{g^\star}^{(2)} = \left[\boldsymbol{s}_{g^\star,K_{g^\star}+1} \cdots \boldsymbol{s}_{g^\star,M_{g^\star}}\right] \in \mathbb{C}^{N_{g^\star} \times (M_{g^\star} - K_{g^\star})}$. This is to further aid UE to recognize the optimal narrow beam.

2) UE top-down recognition: Accordingly, UE needs to modify its beam recognition strategy to a top-down recognition one, in which UE firstly selects the best beam group, and then selects the best narrow beam from the selected group.

With fewer transmissions from BS, the correlations of merging coefficients for different beams are higher, which makes it harder to select the candidate optimal beam in each group according to (31). However, we note that the beamforming gain can still be accurately estimated according to (32), which makes it possible to select the optimal beam set based on the reduced received signals. Specifically, for group $\mathcal{F}_g$, BS only activates the first K_g merged beams $\boldsymbol{F}_g\boldsymbol{s}_{g,1}, \cdots, \boldsymbol{F}_g\boldsymbol{s}_{g,K_g}$, and the corresponding received signals at UE are $\boldsymbol{y}_g^{(1)} = \left[y_{g,1} \cdots y_{g,K_g}\right]^{\mathrm{T}}$. The beamforming gain for the optimal beam $\boldsymbol{f}_{g,n_g^\star}$ in group $\mathcal{F}_g$ can be estimated according to

$$\widehat{q}_g^\star = \frac{N_g}{K_g}\left(\boldsymbol{y}_g^{(1)}\right)^{\mathrm{T}}\left(\boldsymbol{S}_g^{(1)}(n_g^\star,:)\right)^{\mathrm{H}}, \tag{34}$$

Algorithm 2 Proposed ADS-HMBT scheme

Input: G, $\{\mathcal{F}_g\}_{g=1}^G$, $\{N_g\}_{g=1}^G$, $\{K_g\}_{g=1}^G$, $\{M_g\}_{g=1}^G$, $\{\boldsymbol{S}_g\}_{g=1}^G$;
Output: The optimal beam index $\{g^\star, n_{g^\star}^\star\}$;

```
for g = 1 to G do
    for k = 1 to K_g do
        BS transmits merged beam f̄_{g,k} = F_g S_g(:,k);
        UE receives signal y_{g,k};
    end for
    UE computes beamforming gain q̂_g^⋆ for F_g using (34);
end for
UE determines optimal group index g^⋆ using (36);
UE feeds back g^⋆ to BS;
for k = K_{g^⋆} + 1 to M_{g^⋆} do
    BS transmits merged beam f̄_{g^⋆,k} = F_{g^⋆} S_{g^⋆}(:,k),
    UE receives signal y_{g^⋆,k};
end for
UE determines optimal narrow beam using (37) and feeds back n_{g^⋆}^⋆ to BS;
BS uses f_{g^⋆,n_{g^⋆}^⋆} for downlink transmission;
```

where

$$n_g^\star = \arg\max_{1 \leq n_g \leq N_g} \left|\left(\boldsymbol{y}_g^{(1)}\right)^{\mathrm{T}}\left(\boldsymbol{S}_g^{(1)}(n_g,:)\right)^{\mathrm{H}}\right|^2. \tag{35}$$

The index of the optimal group can then be recognized as

$$g^\star = \arg\max_{1 \leq g \leq G} \left|\widehat{q}_g^\star\right|. \tag{36}$$

UE feeds back this optimal group index $g^\star$ to BS. After receiving the optimal group index from UE, BS activates the remaining $(M_{g^*} - K_{g^*})$ merged beams of the optimal group $\mathcal{F}_{g^\star}$ with the MCM $\boldsymbol{S}_{g^\star}^{(2)}$. This help UE to obtain the further $(M_{g^*} - K_{g^*})$ measurements $\boldsymbol{y}_{g^\star}^{(2)} = \left[y_{g^\star,K_{g^\star}+1} \cdots y_{g^\star,M_{g^\star}}\right]^{\mathrm{T}}$. With more merged beams measurements, the correlations of merging coefficients for different beams are decreased. Hence, with the full $M_{g^\star}$ measurement $\boldsymbol{y}_{g^\star}$, UE recognizes the optimal narrow beam according to

$$n_{g^\star}^\star = \arg\max_{1 \leq n_{g^\star} \leq N_{g^\star}} \left|\boldsymbol{y}_{g^\star}^{\mathrm{T}}\left(\boldsymbol{S}_{g^\star}\left(n_{g^\star},:\right)\right)^{\mathrm{H}}\right|^2. \tag{37}$$

3) Discussions: The proposed ADS-HMBT scheme is summarized in **Algorithm 2**, where it can be seen that only $\sum_{g=1}^{G} K_g + \left(M_{g^\star} - K_{g^\star}\right)$ transmissions and two feedback are needed. The computational complexity of the ADS-HMBT is on the order of $\mathsf{O}\left(\sum_{g=1}^{G} N_g K_g + N_{g^\star}\left(M_{g^\star} - K_{g^\star}\right)\right)$. Therefore, by utilizing one more UE feedback information to perform beam training, the ADS-HMBT further reduces the beam training overhead over the ADS-HMBT. Therefore, the ADS-MBT scheme is more suitable for delay-sensitive scenarios, while the ADS-HMBT scheme is more suitable for delay-tolerant scenarios. BS can choose an appropriate scheme according to different situations.

IV. Performance Analysis

Beam misalignment can occur due to multipath and noise. This section provides theoretic beam misalignment probability

analysis on the proposed ADS-MBT scheme. Since the analysis for the ADS-MBT can also be applied to the ADS-HMBT, it is sufficient to just consider the ADS-MBT. In our ADS-MBT scheme, there are two kinds of misalignment, namely, group misrecognition and intra-group beam misrecognition.

A. Group Misrecognition Probability

Group misrecognition happens when the estimated beamforming gain of a non-optimal group $\mathcal{F}_g$ is higher than that of the optimal group $\mathcal{F}_{g^\star}$. Denote a group misrecognition event as $\mathcal{F}_{g^\star} \to \mathcal{F}_g$. Then

$$\mathcal{F}_{g^\star} \to \mathcal{F}_g : \ \exists g \neq g^\star, \ \text{s.t.} \ \left|\widehat{q}_g^\star\right| > \left|\widehat{q}_{g^\star}^\star\right|. \tag{38}$$

By combining (26) and (32), the beamforming gain estimation can be rewritten as

$$\begin{aligned}\widehat{q}_g^\star = \sqrt{\beta} q_{g,n_g^\star} + \sqrt{\beta} \sum_{n_g \neq n_g^\star} q_{g,n_g} \frac{\boldsymbol{S}_g(n_g,:)\big(\boldsymbol{S}_g(n_g^\star,:)\big)^{\mathrm{H}}}{\left|\boldsymbol{S}_g(n_g^\star,:)\big(\boldsymbol{S}_g(n_g^\star,:)\big)^{\mathrm{H}}\right|} \\ + \frac{\boldsymbol{n}^{\mathrm{T}}\big(\boldsymbol{S}_g(n_g^\star,:)\big)^{\mathrm{H}}}{\left|\boldsymbol{S}_g(n_g^\star,:)\big(\boldsymbol{S}_g(n_g^\star,:)\big)^{\mathrm{H}}\right|},\end{aligned} \tag{39}$$

wherein the second term is the interference and is denoted as I_g, while the third term is the noise and is denoted as Ξ_g.

As for I_g, we mainly consider the interferences from neighboring beams, as they typically are the strongest. Hence, for $n_g = n_g^\star \pm 1$, we have

$$q_{g,n_g^\star \pm 1} = \sqrt{N_T} \sum_{p=0}^{P} \gamma_p \boldsymbol{a}_T^{\mathrm{H}}\big(\theta_p^{(AoD)}\big) \boldsymbol{a}_T\big(\theta_{g,n_g^\star \pm 1}^T\big). \tag{40}$$

Since $\boldsymbol{f}_{g,n_g^\star}$ has high beamforming gain, it is reasonable to assume that $\theta_0^{(AoD)}$ is close to $\theta_{g,n_g^\star}^T$. By defining

$$\Delta(g, n_g, i) = \cos\big(\theta_{g,n_g}^T\big) - \cos\big(\theta_i^{(AoD)}\big), \ 1 \leq i \leq P, \tag{41}$$

we have

$$\begin{aligned}\boldsymbol{a}_T^{\mathrm{H}}\big(\theta_p^{(AoD)}\big)\boldsymbol{a}_T\big(\theta_{g,n_g^\star \pm 1}^T\big) &= \frac{\sin\left(\frac{\pi N_T \Delta(g,n_g^\star,i)}{2} + \pi G\right)}{N_T \sin\left(\frac{\pi \Delta(g,n_g,i)}{2}\right)} \\ &\approx \frac{\sin\left(\frac{\pi N_T \Delta(g,n_g^\star,i)}{2} + \pi G\right)}{N_T \sin\left(\frac{\pi G}{N_T}\right)}.\end{aligned} \tag{42}$$

When N_T is large, (42) approximately follows the uniform distribution in the interval $\left[-\frac{1}{N_T \sin\left(\frac{\pi G}{N_T}\right)}, \ \frac{1}{N_T \sin\left(\frac{\pi G}{N_T}\right)}\right]$, denoted as $\mathcal{U}\left(-\frac{1}{N_T \sin\left(\frac{\pi G}{N_T}\right)}, \ \frac{1}{N_T \sin\left(\frac{\pi G}{N_T}\right)}\right)$. Now use the approximation (42) in (40) and consider all the P paths. Since P is typically large, by the central limit theorem, $q_{g,n_g^\star \pm 1}$ follows approximately the complex Gaussian distribution with zero mean and the power

$$\sigma^2_{q_{g,n_g^\star \pm 1}} = \frac{1}{3N_T} \frac{\sum_{p=0}^{P} |\gamma_p^2|}{\sin^2\left(\frac{\pi G}{N_T}\right)}, \tag{43}$$

that is,

$$q_{g,n_g^\star \pm 1} \sim \mathcal{CN}\left(0, \sigma^2_{q_{g,n_g^\star \pm 1}}\right). \tag{44}$$

Let $\delta(\boldsymbol{S}_g)$ be the maximum interference coefficient of $\boldsymbol{S}_g$, i.e.,

$$\delta(\boldsymbol{S}_g) = \max_{\forall i,j,\, i \neq j} \left| \frac{\boldsymbol{S}_g(i,:)\big(\boldsymbol{S}_g(j,:)\big)^{\mathrm{H}}}{\boldsymbol{S}_g(i,:)\big(\boldsymbol{S}_g(i,:)\big)^{\mathrm{H}}} \right|. \tag{45}$$

Then the interference $I_g \sim \mathcal{CN}\left(0, \sigma^2_{I_g}\right)$ with the power

$$\sigma^2_{I_g} = \frac{\beta N_e \delta^2(\boldsymbol{S}_g)}{3N_T} \frac{\sum_{p=0}^{P} |\gamma_p^2|}{\sin^2\left(\frac{\pi G}{N_T}\right)}, \tag{46}$$

where

$$N_e = \frac{1}{\delta^2(\boldsymbol{S}_g)} \sum_{k=\pm 1} \left(\frac{\left|\boldsymbol{S}_g\big(n_g^\star + k,:\big)\big(\boldsymbol{S}_g\big(n_g^\star,:\big)\big)^{\mathrm{H}}\right|}{\left|\boldsymbol{S}_g\big(n_g^\star,:\big)\big(\boldsymbol{S}_g\big(n_g^\star,:\big)\big)^{\mathrm{H}}\right|} \right)^2. \tag{47}$$

Clearly, the noise term $\Xi_g \sim \mathcal{CN}\big(0, \sigma^2_{\Xi_g}\big)$ with the power

$$\sigma^2_{\Xi_g} = \frac{\sigma_n^2}{\left\|\boldsymbol{S}_g(n_g^\star,:)\right\|_2^2}. \tag{48}$$

Let $\sigma^2_{\mathrm{in},g}$ be the power of interference plus noise, i.e., $\sigma^2_{\mathrm{in},g} = \sigma^2_{I_g} + \sigma^2_{\Xi_g}$. Due to high beamforming gain provided by large-scale array, we have $\sqrt{\beta}\left|q_{g,n_g^\star}\right| \gg \sigma_{\mathrm{in},g}$, and consequently the beamforming gain $\left|\widehat{q}_g^\star\right|$ follows the real Gaussian distribution with mean $\sqrt{\beta}\left|q_{g,n_g^\star}\right|$ and variance $\frac{\sigma^2_{\mathrm{in},g}}{2}$, that is,

$$\left|\widehat{q}_g^\star\right| \sim \mathcal{N}\left(\sqrt{\beta}\left|q_{g,n_g^\star}\right|, \frac{\sigma^2_{\mathrm{in},g}}{2}\right). \tag{49}$$

Therefore, the probability of misrecognizing $\mathcal{F}_g$ as the optimal group is given by

$$P_e\big(\mathcal{F}_{g^\star} \to \mathcal{F}_g\big) \approx Q\left(\frac{\sqrt{\beta}\left(\left|q_{g^\star,n_{g^\star}^\star}\right| - \left|q_{g,n_g^\star}\right|\right)}{\sqrt{\frac{\sigma^2_{\mathrm{in},g^\star} + \sigma^2_{\mathrm{in},g}}{2}}} \right), \tag{50}$$

where $Q(\cdot)$ is the complementary cumulative function of Gaussian distribution. The total group misrecognition probability is

$$P_e\big(\mathcal{F}_{g^\star}\big) \leq \sum_{g \neq g^\star} P_e\big(\mathcal{F}_{g^\star} \to \mathcal{F}_g\big). \tag{51}$$

B. Intra-group Beam Misrecognition Probability

Within the group $\mathcal{F}_g$, UE may misrecognize the optimal beam due to sidelobe effect. Intra-group beam misrecognition happens when the estimated beamforming gain of another beam is higher than that of the optimal beam, i.e.,

$$\boldsymbol{f}_{g,n_g^\star} \to \boldsymbol{f}_{g,n_g} : \ \exists n_g \neq n_g^\star, \ \text{s.t.} \ \left|\widehat{q}_{g,n_g}\right| > \left|\widehat{q}_{g,n_g^\star}\right|. \tag{52}$$

Since the distribution of $\left|\widehat{q}_{g,n_g^\star}\right|$ is given in (49), we only need to concentrate on $\left|\widehat{q}_{g,n_g}\right|$, $n_g \neq n_g^\star$. Similar to (39) to (44), the beamforming gain estimation for q_{g,n_g} can be expressed as

$$\widehat{q}_{g,n_g} \approx \sqrt{\beta} q_{g,n_g^\star} \frac{\boldsymbol{S}_g(n_g,:)\big(\boldsymbol{S}_g(n_g^\star,:)\big)^{\mathrm{H}}}{\left|\boldsymbol{S}_g(n_g,:)\big(\boldsymbol{S}_g(n_g,:)\big)^{\mathrm{H}}\right|} + I_g^{'} + \Xi_g^{'}, \tag{53}$$

where $I_g^{'}$ is the interference term, and $\Xi_g^{'}$ is the noise term.

Obviously, the noise $\Xi_g^{'} \sim \mathcal{CN}\big(0, \sigma^2_{\Xi_g^{'}}\big)$ with the power $\sigma^2_{\Xi_g^{'}} = \frac{\sigma_n^2}{\left\|\boldsymbol{S}_g(n_g,:)\right\|_2^2}$. As for $I_g^{'}$, due to the strong directivity

of beam, we can only consider the spatial neighboring beams $(n_g = n_g^\star \pm 1)$. By defining

$$\boldsymbol{R}(i,j) = \frac{\boldsymbol{S}_g(i,:)\big(\boldsymbol{S}_g(j,:)\big)^{\mathrm{H}}}{\Big|\boldsymbol{S}_g(i,:)\big(\boldsymbol{S}_g(i,:)\big)^{\mathrm{H}}\Big|}, \tag{54}$$

we have

$$I_g^{'} = \boldsymbol{R}\big(n_g, n_g^\star - 1\big) q_{g,n_g^\star - 1} + \boldsymbol{R}\big(n_g, n_g^\star + 1\big) q_{g,n_g^\star + 1}. \tag{55}$$

Noting (43) and (44) leads to $I_g^{'} \sim \mathcal{CN}\Big(0, \sigma^2_{I_g^{'}}\Big)$ with the power

$$\sigma^2_{I_g^{'}} = \frac{\beta \sum_{k=\pm 1} \big|\boldsymbol{R}\big(n_g, n_g^\star + k\big)\big|^2}{3N_T} \frac{\sum_{p=0}^{P} \big|\gamma_p^2\big|}{\sin^2\left(\frac{\pi G}{N_T}\right)}. \tag{56}$$

Thus the distribution of the amplitude $\big|\widehat{q}_{g,n_g}\big|$ is approximately Gaussian, specifically,

$$\big|\widehat{q}_{g,n_g}\big| \sim \mathcal{N}\left(\sqrt{\beta}\big|\boldsymbol{R}\big(n_g, n_g^\star\big)\big|\big|q_{g,n_g^\star}\big|, \frac{\sigma^2_{\mathrm{in}^{'},g}}{2}\right), \tag{57}$$

wherein $\sigma^2_{\mathrm{in}^{'},g} = \sigma^2_{I_g^{'}} + \sigma^2_{\Xi_g^{'}}$. The pairwise misrecognition probability is given by

$$P_e\big(\boldsymbol{f}_{g,n_g^\star} \to \boldsymbol{f}_{g,n_g}\big) \approx Q\left(\frac{\sqrt{\beta}\big|q_{g,n_g^\star}\big|\big(1 - \big|\boldsymbol{R}(n_g, n_g^\star)\big|\big)}{\sqrt{\frac{\sigma^2_{\mathrm{in},g} + \sigma^2_{\mathrm{in}^{'},g}}{2}}}\right), \tag{58}$$

and the total intra-beam misrecognition probability is given by

$$P_e\big(\boldsymbol{f}_{g,n_g^\star}\big) \le \sum_{n_g \neq n_g^\star} P_e\big(\boldsymbol{f}_{g,n_g^\star} \to \boldsymbol{f}_{g,n_g}\big). \tag{59}$$

C. Discussions

Including the group and intra-group beam misrecognition, the total misrecognition probability is

$$\begin{aligned} P_e =& 1 - \big(1 - P_e\big(\mathcal{F}_{g^\star}\big)\big)\big(1 - P_e\big(\boldsymbol{f}_{g^\star,n_{g^\star}^\star}\big)\big) \\ \approx& P_e\big(\mathcal{F}_{g^\star}\big) + P_e\big(\boldsymbol{f}_{g^\star,n_{g^\star}^\star}\big). \end{aligned} \tag{60}$$

The above two misrecognition influence the performance of the proposed scheme differently. When group misrecognition $\mathcal{F}_{g^\star} \to \mathcal{F}_g$ occurs, UE selects a suboptimal[1] beam $\boldsymbol{f}_{g,n_g^\star}$ with beamforming gain $q_{g,n_g^\star}$, which will not cause severe beamforming gain degradation. However, when intra-group beam recognition happens, UE will select a beam with almost no gain, which causes severe performance degradation. From (58) and (59), it can be seen that the MCM $\boldsymbol{S}_g$ significantly influences the performance of the scheme.

The aforementioned beam misrecognition analysis for the ADS-MBT scheme is equally applicable to the ADS-HMBT scheme. Specifically, the first part of the ADS-HMBT selects the optimal group, and to analyze the group misrecognition probability of the ADS-HMBT, we only need to replace $\boldsymbol{S}_g$ with $\boldsymbol{S}_g^{(1)}$ in Subsection IV-A. The second part of the ADS-HMBT selects the optimal beam within the chosen optimal group, and we can analyze the intra-optimal-group beam misrecognition probability of the ADS-HMBT by applying the results of Subsection IV-B to the selected optimal group $\mathcal{F}_{g^\star}$.

Remark 1: For multi-antenna UE, the interactive search also causes beam misalignment. Further taking into account UE side beam training, the total misalignment probability of the beam training for multi-antenna UE can be expressed as

$$\begin{aligned} P_e \approx& P_{e,BS} + P_{e,ia} + P_{e,UE} \\ \approx& P_e\big(\mathcal{F}_{g^\star}\big) + P_e\big(\boldsymbol{f}_{g^\star,n_{g^\star}^\star}\big) + P_{e,ia} + P_{e,UE}, \end{aligned} \tag{61}$$

where $P_{e,BS}$ is the misalignment probability for BS beam, which is given in (60), $P_{e,ia}$ is the misalignment probability by interactive search, i.e., the probability that interactive search ((12) and (13)) cannot obtain the best beam pair, and $P_{e,UE}$ is the misalignment probability for UE beam. Define

$$P_{e,ma} = P_{e,ia} + P_{e,UE} \tag{62}$$

as the beam misalignment probability introduced by multi-antenna UE. Obviosuly, as the number of UE antennas increases, the size of UE codebook increases, which increases the probability of selecting the neighboring suboptimal beam, and this increases $P_{e,ma}$. In addition, as the effect of channel multi-path increases, $P_{e,ma}$ also increases.

V. Merging Coefficient Matrix Design

We now discuss how to optimize the MCM design so as to improve the beam misalignment performance. According to the analysis presented in the previous section, specifically, (38) to (60), our aim is to find an $N_g \times M_g$ matrix $\boldsymbol{S}_g$ that minimizes $\delta(\boldsymbol{S}_g)$. If $M_g \ge N_g$, $\boldsymbol{S}_g$ can always be designed with orthogonal rows to minimize $\delta(\boldsymbol{S}_g)$, i.e., $\boldsymbol{S}_g \boldsymbol{S}_g^{\mathrm{H}} = \boldsymbol{I}_{N_g}$ and $\delta(\boldsymbol{S}_g) = 0$. However, in this case, training overhead cannot be reduced compared to the exhaustive search.

To reduce the training overhead, in our design, we set $M_g < N_g$, which implies that the rows in $\boldsymbol{S}_g$ cannot be made completely orthogonal to each other. Intuitively, by increasing M_g, the row vectors of $\boldsymbol{S}_g$ can be selected to be closer to orthogonal and the power accumulated for each narrow beam increases, hence improving the performance. By contrast, increasing N_g makes $\boldsymbol{S}_g$ further away from the orthogonality and decreases the power accumulated for each narrow beam, which degrades the performance. Our simulation results (Figs. 7 and 9 in Section VI) show that $M_g/N_g \approx 1/2$ achieves good performance between training overhead and beamforming gain, while the parameter configuration $(M_g, N_g) = (3, 8)$ suffers serious performance degradation.

Given M_g and N_g, we can carefully design the MCM to make $\boldsymbol{S}_g$ as close to orthogonal as possible in order to improve the performance. We omit the subscript g for notational simplicity in our discussion. The design of MCM can be formulated as the following optimization problem

$$\begin{aligned} \min_{\boldsymbol{S} \in \mathbb{C}^{N \times M}} \quad & \delta(\boldsymbol{S}), \\ \text{s.t.} \quad & \|\boldsymbol{S}(:,m)\|_2 = 1, \ 1 \le m \le M. \end{aligned} \tag{63}$$

[1]A beam is 'suboptimal' if it is not the optimal beam but still has high beamforming gain.

This problem is equivalent to Grassmannian line packing and it has been studied by many researchers. The Rankin lower bound [32], [33] is given by

$$\delta(\boldsymbol{S}) \geq \sqrt{1-\frac{(M-1)N}{M(N-1)}}. \tag{64}$$

For example, for $M=2$ and $N=4$, an optimal design of MCM reaching the lower bound $\delta(\boldsymbol{S})=\frac{1}{\sqrt{3}}$ is given by

$$\boldsymbol{S}=\frac{1}{\sqrt{2}}\begin{bmatrix}\cos\varphi & \sin\varphi\\ \sin\varphi & j\cos\varphi\\ \cos\varphi & -\sin\varphi\\ \sin\varphi & -j\cos\varphi\end{bmatrix}, \varphi=\arctan\frac{\sqrt{6}-\sqrt{2}}{2}. \tag{65}$$

But for general M and N with $M<N$, the optimal design is hard to solve [34]. To simplify this problem, a popular way is to introduce an equal gain transmission (EGT) constraint [35]–[37], which requires every element of $\boldsymbol{S}$ has the same modulus. This leads to

$$\begin{aligned}&\min_{\boldsymbol{S}\in\mathbb{C}^{N\times M}} \delta(\boldsymbol{S}),\\ &\text{s.t.} \quad |\boldsymbol{S}(n,m)|=\tfrac{1}{\sqrt{N}},\ 1\leq n\leq N, 1\leq m\leq M.\end{aligned} \tag{66}$$

In the following, we introduce a quantized EGT (QEGT) MCM design [38] and a random EGT MCM design.

A. QEGT MCM Design

The authors of [38] propose to choose a set of M columns from the $N\times N$ discrete Fourier transform (DFT) matrix, i.e.,

$$\boldsymbol{S}(n,m)=\frac{1}{\sqrt{N}}e^{\mathrm{j}\frac{2\pi}{N}u_m(n-1)}, 1\leq m\leq M, 1\leq n\leq N, \tag{67}$$

where $0\leq u_1,\cdots,u_M\leq N-1$. The correlation coefficient between the n-th row and the n'-th row is

$$R(n,n')=\frac{1}{M}\sum_{m=1}^{M}e^{\mathrm{j}\frac{2\pi}{N}u_m(n-n')}, \tag{68}$$

which depends only on $\Delta n=(n-n')\bmod N$. The problem is then converted into the following optimization

$$\min_{0\leq u_1,\cdots,u_M\leq N-1}\ \max_{\Delta n=1,\cdots,N-1}\ \frac{1}{M}\left|\sum_{m=1}^{M}e^{\mathrm{j}\frac{2\pi}{N}u_m\Delta n}\right|. \tag{69}$$

To solve for the optimal $u_1^\star,\cdots,u_M^\star$, one can use an exhaustive search for small M and N, or performs a random search for large M and N to get a satisfactory solution. Generally, M and N are not large, and $\binom{M}{N}$ exhaustive search complexity is acceptable. For $M=2$ and $N=4$, a QEGT MCM design is

$$\boldsymbol{S}=\frac{1}{2}\begin{bmatrix}1 & 1\\ 1 & j\\ 1 & -1\\ 1 & -j\end{bmatrix},\ \delta(\boldsymbol{S})=\frac{1}{\sqrt{2}}. \tag{70}$$

Remark 2: When combining this QEGT MCM design with the merging principle (18), the beam vector $\bar{\boldsymbol{f}}$ has N_T/N elements with gain $\frac{1}{\sqrt{N_T/N}}$, and the rest of the elements are exactly 0. Hence BS can shape the merged beam with single RF chain by only activating the corresponding antennas.

B. Random EGT MCM Design

For large M and N, we show that random matrix is a feasible design [39]. We choose $\boldsymbol{S}(n,m)=\frac{1}{\sqrt{N}}e^{\mathrm{j}\theta_{n,m}}$, where $\forall n,m$, $\theta_{n,m}$ follow the independent uniform distribution in the interval $[0,\ 2\pi)$, i.e., $\theta_{n,m}\sim\mathcal{U}(0,\ 2\pi)$. The correlation coefficient between the n-th row and n'-th row is

$$R(n,n')=\frac{1}{M}\sum_{m=1}^{M}e^{\mathrm{j}(\theta_{n,m}-\theta_{n',m})}, \tag{71}$$

For large M, using the central limit theorem, we have

$$R(n,n')\overset{M\to\infty}{\sim}\mathcal{CN}\left(0,\ \frac{1}{M}\right). \tag{72}$$

Let $z=|R(n,n')|$. The cumulative distribution function of z can be expressed as

$$F(z)=1-e^{-Mz^2}, z\geq 0. \tag{73}$$

For a randomly generated matrix, the probability of $|R(n,n')|<\xi$, $\forall 1\leq n<n'\leq N$, is given by

$$\begin{aligned}\Pr(\delta(\boldsymbol{S})<\xi)=&1-\Pr\left(\bigcup\nolimits_{1\leq n<n'\leq N}|R(n,n')|>\xi\right)\\ \leq&1-\sum_{1\leq n<n'\leq N}\Pr(|R(n,n')|>\xi)\\ ==&1-\frac{N(N-1)}{2}e^{-M\xi^2}.\end{aligned} \tag{74}$$

For fixed ξ, when $M,N\to\infty$ with fixed M/N, $\delta(\boldsymbol{S})<\xi$ holds with probability $\Pr\to 1$. One can randomly generate several matrices until the condition $\delta(\boldsymbol{S})<\xi$ is met.

C. A View of Beam Splitting

From the viewpoint of beam splitting, we can provide some insights on our proposed ADS-MBT scheme. From (29)-(33), we see that our ADS-MBT scheme estimates beamforming gain for $\boldsymbol{f}_{g,n_g}$ as $\widehat{q}_{g,n_g}$, $\forall g,n_g$, and selects the beam with the highest estimated beamforming gain. For a beam $\boldsymbol{f}_{g,n_g}\in\mathcal{F}_g$, we simplify $\boldsymbol{f}_{g,n_g}$ as $\boldsymbol{f}_n$ by omitting the subscript g again. Combining (24) to (26) and (29) to (33), the beamforming gain estimation can be expressed as

$$\begin{aligned}\widehat{q}_n=&\frac{\boldsymbol{y}^{\mathrm{T}}(\boldsymbol{S}(n,:))^{\mathrm{H}}}{\boldsymbol{S}(n,:)(\boldsymbol{S}(n,:))^{\mathrm{H}}}\\ =&\sqrt{\beta}\boldsymbol{h}^{\mathrm{H}}\frac{\boldsymbol{F}\boldsymbol{S}(\boldsymbol{S}(n,:))^{\mathrm{H}}}{\boldsymbol{S}(n,:)(\boldsymbol{S}(n,:))^{\mathrm{H}}}+\frac{\boldsymbol{n}^{\mathrm{T}}(\boldsymbol{S}(n,:))^{\mathrm{H}}}{\boldsymbol{S}(n,:)(\boldsymbol{S}(n,:))^{\mathrm{H}}}\\ =&\sqrt{\beta}\boldsymbol{h}^{\mathrm{H}}\boldsymbol{f}_{\mathrm{eq},n}+n_{\mathrm{eq},n},\ 1\leq n\leq N,\end{aligned} \tag{75}$$

in which $\boldsymbol{f}_{\mathrm{eq},n}=\frac{\boldsymbol{F}\boldsymbol{S}(\boldsymbol{S}(n,:))^{\mathrm{H}}}{\boldsymbol{S}(n,:)(\boldsymbol{S}(n,:))^{\mathrm{H}}}$ can be regarded as the n-th equivalent transmission beam.

For UE, M repeatedly transmissions of the merged beams are equivalent to N transmissions of the equivalent beams. UE uses the beamforming gain of $\boldsymbol{f}_{\mathrm{eq},n}$ as the approximation to that of $\boldsymbol{f}_n$. As for the equivalent beam $\boldsymbol{f}_{\mathrm{eq},n}$, we have

$$\boldsymbol{f}_{\mathrm{eq},n}=\boldsymbol{f}_n+\sum_{n'\neq n}\frac{\boldsymbol{S}(n',:)(\boldsymbol{S}(n,:))^{\mathrm{H}}}{\boldsymbol{S}(n,:)(\boldsymbol{S}(n,:))^{\mathrm{H}}}\boldsymbol{f}_{n'}, \tag{76}$$

where the second term can be regarded as the interference sidelobes. Hence $\delta(\boldsymbol{S})$ represents the upper bound of the strength of sidelobes. Small $\delta(\boldsymbol{S})$ means small sidelobe interference.

Fig. 5 (a) shows a beam group example consisting 4 beams, and Fig. 5 (b) and (c) show the merged beam patterns and split equivalent beam patterns. In Fig. 5 (b) and (c), 4 beams are merged to $\tilde{\boldsymbol{f}}_1$ and $\tilde{\boldsymbol{f}}_2$ according to the two different MCM $\boldsymbol{S}$. From the viewpoint of beam splitting, the merged beams are split into 4 equivalent beams. Compared with the original beam patterns shown in Fig. 5 (a), an equivalent beam has one original beam pattern as the mainlobe and the other patterns as sidelobes, and the correlation coefficient $\frac{\boldsymbol{S}(n',:)(\boldsymbol{S}(n,:))^{\mathrm{H}}}{\boldsymbol{S}(n,:)(\boldsymbol{S}(n,:))^{\mathrm{H}}}$ represents the strength of sidelobe. For different MCM $\boldsymbol{S}$, sidelobe patterns may be different. In Fig. 5 (b) with the MCM (65), each equivalent beam has 3 small sidelobes, while in Fig. 5 (c) with the MCM (65), each equivalent beam has only 2 but larger sidelobes.

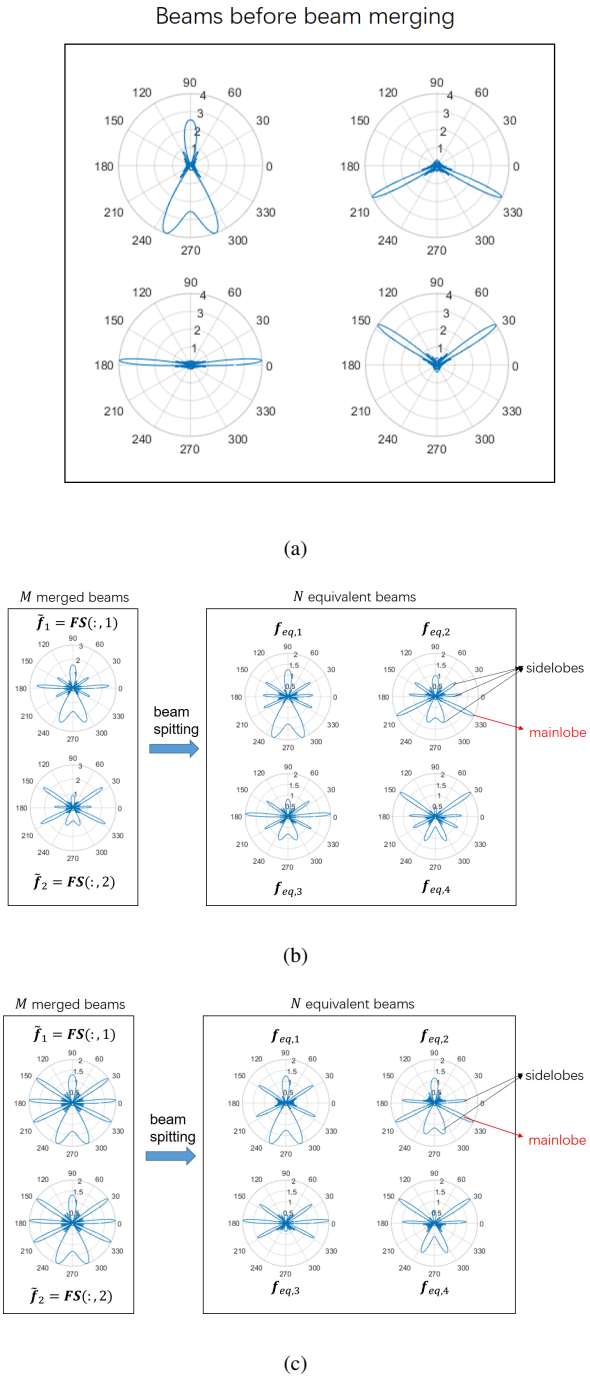


Fig. 5. Beam splitting pattern examples with different MCM $\boldsymbol{S}$. (a) Beam patterns before beam merging. (b) Beam splitting pattern for MCM (65). (c) Beam splitting pattern for MCM (70). From the viewpoint of beam splitting, M merged beam transmissions are split into N beams. Splitting beams have one main lobe and several side lobes.

VI. Numeric Simulations

In our simulation system, BS employs half-wavelength spaced ULA and uses the DFT codebook based beamforming with the number of antennas $N_T = 64$. For the main part of the simulation study, UE is equipped with single antenna ($N_R = 1$), As for the channel vector $\boldsymbol{h}$, its complex gains $\gamma_0 \sim \mathcal{CN}(0, \sigma^2_{\mathrm{LoS}})$ and $\gamma_p \sim \mathcal{CN}(0, \sigma^2_{\mathrm{NLoS}})$, $1 \le p \le P$. The power ratio of the LoS path to the NLoS paths is denoted as $\rho = \sigma^2_{\mathrm{LoS}}/\sigma^2_{\mathrm{NLoS}}$, and the angular spread is set to $\theta_{AS} = 15°$. The system's signal to noise ratio (SNR) is defined as $\mathrm{SNR} = \beta/\sigma^2_n$. In all the simulations, we adopt the grouping principle (18) and obtain the MCM $\boldsymbol{S}$ from the QEGT design of Subection V-A. We present the numeric simulation results to evaluate the performance of our ADS-MBT and ADS-HMBT. Additionally, we also validate our proposed schemes for the multi-antenna UE case in the simulation study.

A. ADS-MBT Performance Evaluation

We compare the performance of our ADS-MBT scheme with the exhaustive search (ES) scheme and the calibrate beam training scheme [14]. The default beam merging parameters are set to $G = N_T/4$, $N_g = N = 4$ and $M_g = M = 2$. All these schemes require single feedback in beam training. The ES scheme requires N_T time slots to sweep all the beams, while the scheme of [14] and our ADS-MBT scheme require only $\frac{N_T}{2}$ time slots. In terms of computational complexity, the ES compares the receiving powers of the N_T beams, and its complexity is on the order of $\mathsf{O}(N_T)$. The scheme [14] compares the receiving powers of $N_T/2$ wide beams and computes the power differences between the optimal wide beam and its adjacent beams, and hence its complexity is $\mathsf{O}(N_T/2+2)$. As mentioned in Subsection III-B, the complexity of our ADS-MBT is $\mathsf{O}(2N_T)$. All these schemes have linear complexity in N_T. The performance of these methods are evaluated, in terms of beam misalignment probability and beamforming gain.

Fig. 6 (a) compares the beam misalignment probability performance of the ADS-MBT under different multipath conditions with those of other schemes. The theoretical results of the ADS-MBT are calculated according to the analytical results of Section IV, while the simulation results are obtained by Monte Carlo method. From Fig. 6 (a), it can be seen that the theoretical results closely match the simulation results, indicating that the analysis of Section IV gives a very good approximation to the true beam misalignment probability. As expected, the ES achieves the lower beam misalignment probability, since it attains the true optimal solution. Our ADS-MBT outperforms the scheme [14] considerably, especially when multipath effect is serious. Observe from Fig. 6 (a) that the performance of our ADS-MBT is close to that of

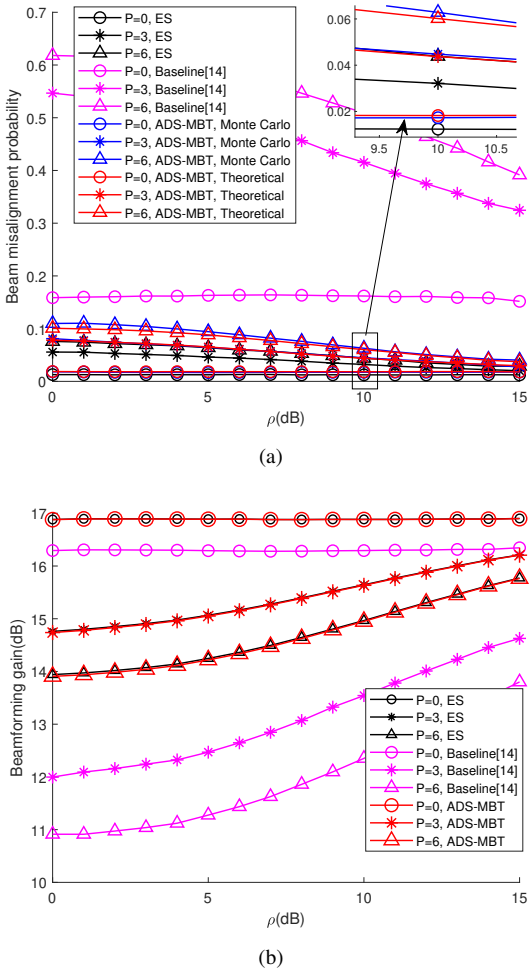


Fig. 6. Performance comparison under different multipath conditions for various schemes with SNR= 10 dB. All these schemes utilize single feedback in beam training. (a) Beam misalignment probability comparison, and (b) Beamforming gain comparison.

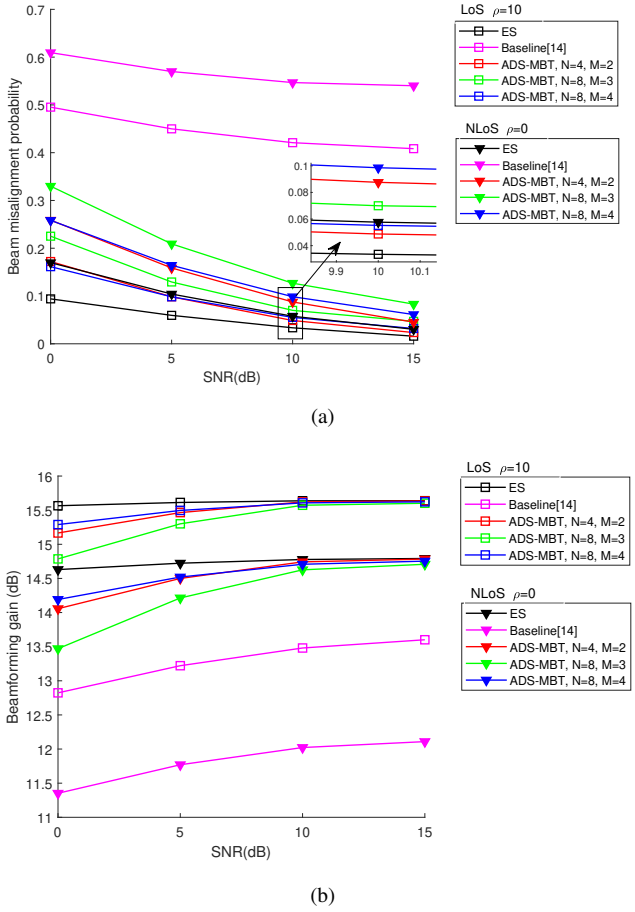


Fig. 7. Performance comparison under different SNR for various schemes with $P = 3$ and two different ρ. All these schemes utilize single feedback in beam training. (a) Beam misalignment probability comparison, and (b) Beamforming gain comparison.

the ES. Also as ρ increases and/or P decreases, the LoS path dominates the beamforming gain, and the performance gap between the ADS-MBT and the ES reduces. As pointed out in Section IV, beam misalignment does not necessarily result in serious performance degradation, and intra-group beam misalignment often results in selecting a suboptimal beam. This is supported by the beamforming gain comparison depicted in Fig. 6 (b), where it is seen that the beamforming gain achieved by our ADS-MBT is very close to the optimal ES scheme. Specifically, our scheme achieves around 99% gain of the ES, under various multi-path conditions.

Fig. 7 (a) and (b) provide the performance comparison at different SNR with the multipath parameters of $P=3$ as well as $\rho=0$ and $\rho=10$. In this experiment, we experience various merging parameters for the ADS-MBT. As can be observed from Fig. 7 (b), our ADS-MBT attains a beamforming gain close to that of the optimal ES scheme, particularly in the high SNR scenario (SNR >10 dB) or when M/N is sufficiently large ($M/N \geq 1/2$). Essentially, when the equivalent receiving power for each narrow beam at UE is sufficiently high, the proposed method can approach the performance of the optimal ES very well. Also observe that our ADS-MBT outperforms the scheme [14] considerably in all the scenarios.

B. ADS-HMBT Performance Evaluation

We compare the performance of our ADS-HMBT scheme with the 2-resolution hierarchical search (two stage search for short) [10] and the subarray-based beam training scheme [21]. For fairness, appropriate parameters are set to ensure that all these schemes use two feedback. The default parameters of the ADS-HMBT are set to $G=N_T/4$, $N_g=N=4$, $M_g=M=2$, and $K_g = 1$, $1 \leq g \leq G$. For the two stage search and subarray-base beam training, each wide beam contains $N = 4$ narrow beams, so that the number of wide beams equals to $G = \frac{N_T}{4}$. For the above parameter settings, the two-stage search and subarray-based search require $G + N = \frac{N_T}{4} + 4$ beam sweeping time slots, while our ADS-HMBT requires $G + K_{g^\star} = \frac{N_T}{4} + 1$ beam sweeping time slots. The computational complexity of the ADS-HMBT is on the order of $\mathsf{O}\left(\sum_{g=1}^{G} N_g K_g + N_{g^\star}\left(M_{g^\star} - K_{g^\star}\right)\right) = \mathsf{O}(N_T + N)$. Both the two-stage search [10] and the subarray-based scheme [21] need to scan all the wide beams and the narrow beams in the optimal wide beam range, and hence their complexity are on the order of $\mathsf{O}(N_T/N + N)$. Clearly, all the three schemes has linear complexity in N_T. Similarly, we evaluate the performance in terms of beam misalignment probability and beamforming gain.

Fig. 8 (a) and (b) compare the beam misalignment prob-

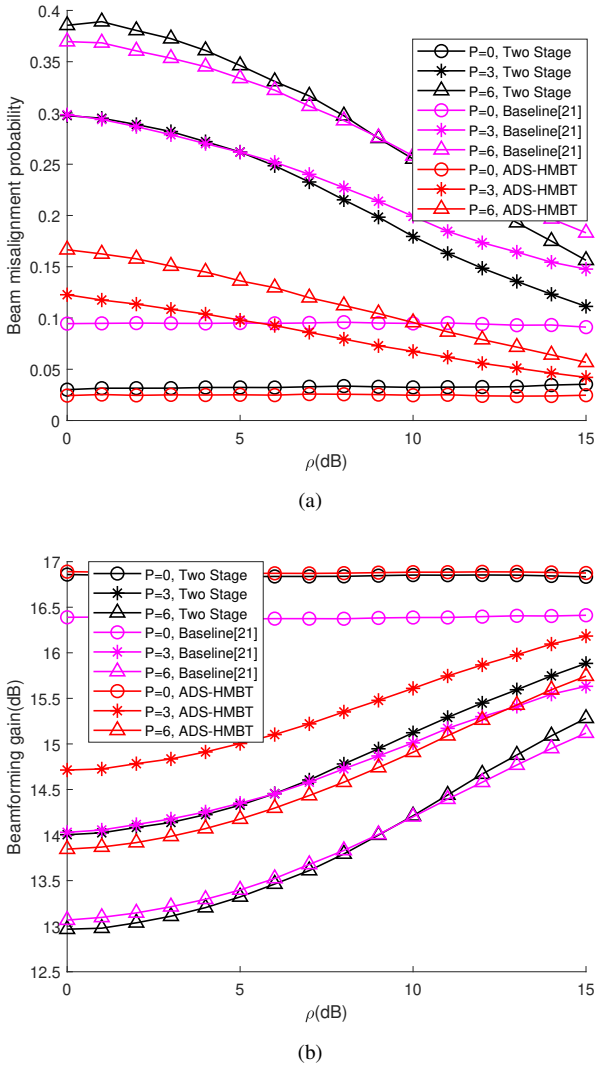


Fig. 8. Performance comparison under different multipath conditions for various schemes with SNR $= 10$ dB. All these schemes use two feedback in beam training. (a) Beam misalignment probability comparison, and (b) Beamforming gain comparison.

ability and beamforming gain performance, respectively, for the schemes under different multipath conditions. The results of Fig. 8 demonstrate that our ADS-HMBT outperforms the two-stage search [10] and the subarray-based benchmark [21], especially when multipath effect is prominent ($\rho \leq 5$dB and $P \geq 3$). More specifically, Fig. 8 (a) shows that our ADS-HMBT reduces the beam misalignment probability by more than 50%, compared to the two benchmark schemes, while Fig. 8 (b) shows that our ADS-HMBT scheme improves the beamforming gain by 0.5 dB to 0.8 dB under various multipath conditions, over the two benchmark schemes. The performance improvement mainly comes from the ADS merging principle. Specifically, in our ADS-HMBT, the optimal merged (wide) beam is more easily selected due to the ADS property (22), and the gain gap of narrow beams in a group is larger, which makes it easier to distinguish the optimal narrow beam from others.

Fig. 9 (a) and (b) depict the beam misalignment probability and beamforming gain performance, respectively, at different SNR for our ADS-HMBT with different emerging parameters

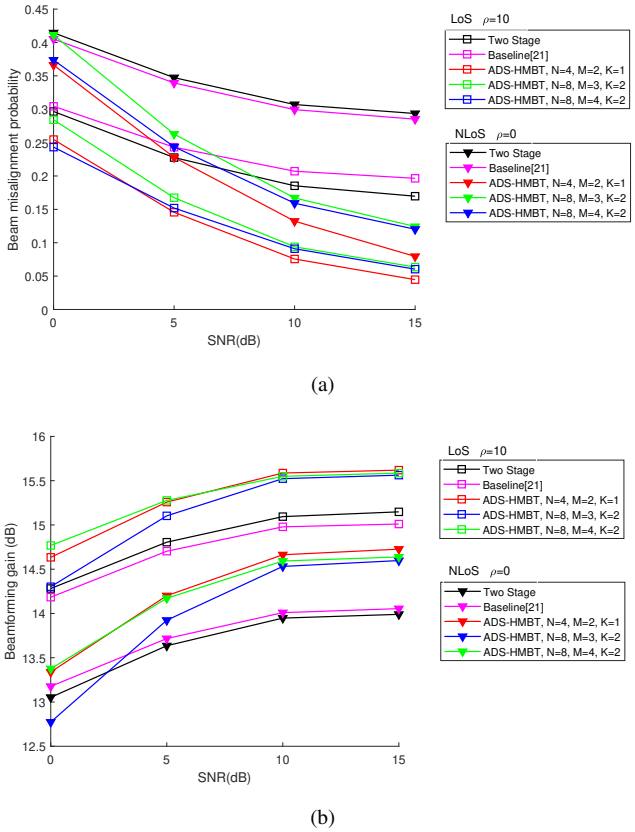


Fig. 9. Performance comparison under different SNR for various schemes with $P = 3$ and two different ρ. All these schemes use two feedback in beam training. (a) Beam misalignment probability comparison, and (b) Beamforming gain comparison.

and the two benchmarks. It can be seen from Fig. 9 (a) that the misalignment probability of our ADS-HMBT decreases rapidly with the increase of SNR. Under both LoS-dominant and multipath scenarios, the ADS-HMBT outperforms the two benchmark schemes. It can be seen from Fig. 9 (b) that the beamforming gain obtained by the ADS-HMBT is higher than those of the two benchmarks when SNR ≥ 4 dB or M/N is sufficient large ($M/N \geq 1/2$). Basically, the ADS-HMBT makes better use of the channel angular domain property, compared to the two-stage search and subarray-based beam training, which enables it to achieve better performance both in training overhead and beamforming gain. The only situation that our ADS-HMBT performs poorer than the two benchmarks is for the merging parameters of $N = 8$, $M = 3$ and $K = 2$ as well as SNR < 4 dB. If M/N is too small, the narrow beams in a group may be difficult to distinguish under poor SNR condition, leading to performance degradation.

C. Beam Training Overhead Comparison

Thirdly we compare the beam training overhead among various beam training schemes in Fig. 10 (a) and (b), in terms of time slots required. The training overhead and computational complexity comparison are summarized in Table I. The ES, the calibrated beam training [14] and our ADS-MBT only require one feedback, while the two stage search [10], the subarray-based beam training [21] and our ADS-HMBT require two feedback. Generally, time slots overhead is proportion to the

TABLE I
THE TRAINING OVERHEAD AND COMPUTATIONAL COMPLEXITY COMPARISON

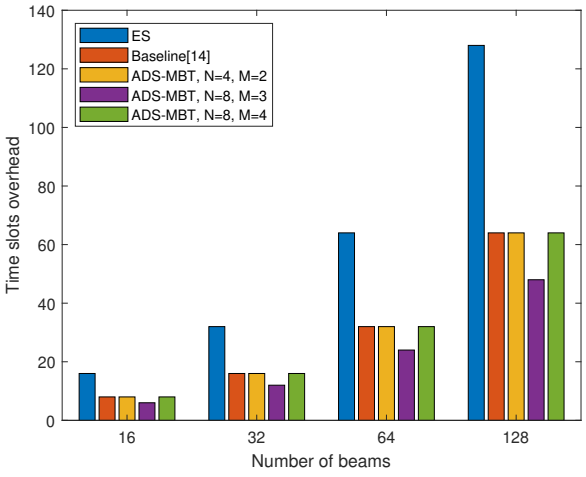
	Exhaustive	**Baseline [14]**	**Hierarchical [10], [21]**	**ADS-MBT**	**ADS-HMBT**
Required transmissions	N_T	$N_T/2$	$N_T/N+N$	MN_T/N	$KN_T/N+(M-K)$
Required feedback	1	1	2	1	2
Computational complexity	$\mathsf{O}(N_T)$	$\mathsf{O}(N_T/2)$	$\mathsf{O}(N_T/N+N)$	$\mathsf{O}(MN_T)$	$\mathsf{O}(KN_T+MN)$

number of beams. It can be seen from Fig. 10 (a) that our ADS-MBT significantly reduces time slot overhead, compared with the optimal ES. Depending on merging parameters, our ADS-MBT imposes the same or lower number of time slots, compared to the calibrate beam training [14]. Compared with the two stage search and the subarray-based beam training [21], Fig. 10 (b) shows that our ADS-HMBT imposes lower time slot overhead. The results also confirm that our ADS-HMBT further reduces time slot overhead over our ADS-MBT, at the expense of an extra feedback.

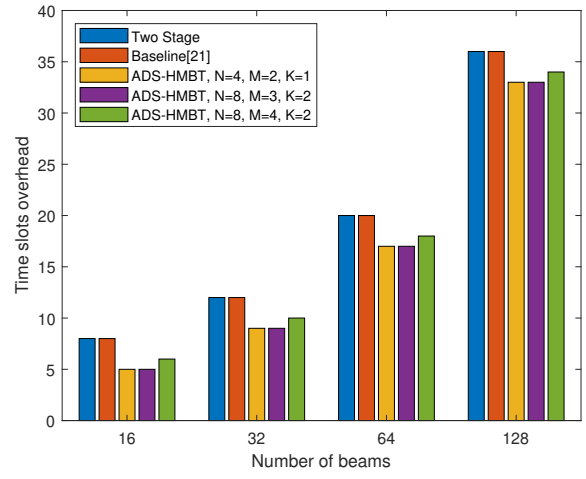
The results of Fig. 10 (a) and (b) indicate that in terms of beam training overhead, our proposed schemes have clearly advantage over the ES and the existing beam training schemes. In terms of beamforming performance, as shown in Subsection VI-A, our ADS-MBT closely approaches the optimal beamforming gain of the ES, and it significantly outperforms the calibrate beam training [14]. As can be seen in Subsection VI-B, our ADS-HMBT scheme significantly improves the achievable beam training performance over the two stage search [10] and the subarray-base beam training [21]. Furthermore, all these schemes compared have computational complexity on the order of N_T. The numerical simulation results therefore demonstrate that our proposed schemes achieve better tradeoff between beaming training overhead and beamforming gain.

D. Performance Comparison for Multi-Antenna UE

Finally, we validate the effectiveness of our proposed schemes for multi-antenna UE. According to IEEE 802.11ad specifications [30], interactive-search is adopted to decou-

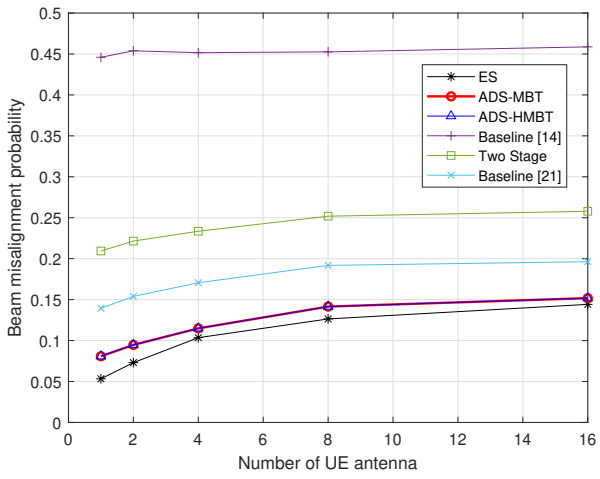


(a)

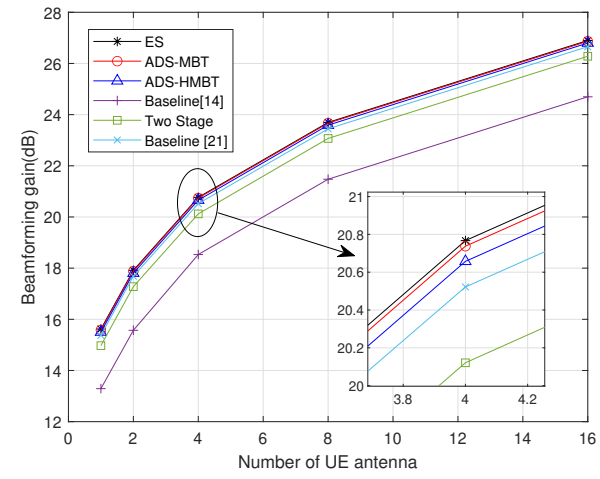


(b)

Fig. 10. Time slot overhead comparison among various beam training schemes. (a) Overhead comparison for one feedback beam training schemes, and (b) Overhead comparison for two feedback beam training schemes.



(a)



(b)

Fig. 11. Performance comparison among various beam training schemes in multi-antenna UE case with SNR $= 5\,$dB and $\rho = 10\,$dB. (a) Beam misalignment probability comparison, and (b) Beamforming gain comparison.

ple UE beamforming and BS beamforming. Our proposed schemes are applied to BS side beam training. The beam merging parameters of our proposed schemes are set to $M=2$, $N=4$ and $K=1$, while SNR is set to $5\,\mathrm{dB}$ and ρ is set to $10\,\mathrm{dB}$. Fig. 11 (a) presents the beam misalignment probability comparison. As the number of UE antenna increases, the size of UE codebook increases, which increases the probability of selecting the neighboring suboptimal beam. Hence, the beam misalignment probability increases with the number of UE antennas. Nevertheless, it can be seen from Fig. 11 (b) that the beamforming gain degradation of our proposed schemes compared to the optimal ES scheme is very small. Specifically, our proposed schemes can still achieve 99% gain of the ES. Moreover, our proposed schemes outperform other benchmark schemes in terms of both misalignment probability and beamforming gain, which demonstrates the effectiveness of our proposed schemes in the multi-antenna UE case.

VII. Conclusions

In this paper, we have proposed the ADS-MBT and ADS-HMBT schemes to support multiple beam simultaneously training with substantially reduced beam training overhead in the one data stream application. We have carried out the detailed theoretical analysis of beam misrecognition for our proposed schemes, which has further guided the design of MCM. Moreover, numerical results have shown the effectiveness and superiority of our proposed beam training schemes. More specifically, like the optimal exhaustive search, our ADS-MBT only requires one feedback but it is capable of achieving almost the same beamforming gain as the exhaustive search with significantly lower beam training overhead. Moreover, compared with the calibrate beam training scheme which has the same beam training overhead, the beamforming gain of our ADS-MBT is significantly better. Our ADS-HMBT exploits two feedback to further reduce the beam training overhead. Compared with the existing two stage search scheme and the subarray-based beam training scheme, which also require two feedback, our ADS-HMBT achieves considerably better beam training performance while imposing lower beam training overhead.

In this work, we focus on the single-stream transmission in angular concentrated channels. Multi-stream transmission is often considered in the rich-scattering environment, where the angular concentration property may not hold. If there are multiple high-gain beams in different groups, our proposed schemes still work. However, if there are multiple high-gain beams in one group, the beamforming gain estimation accuracy may be affected. Hence, our proposed schemes may suffer from performance degradation. Our future work will investigate the robust extension of our proposed schemes to the multi-stream transmission.

Appendix
UE feed back procedure in initial access

In the initial access beam training procedure, the purpose of UE feedback is to inform BS the optimal beam selected by UE. When UE feeds back, BS does not know which beamforming vector should be used to receive the feedback. In other words, uplink has not been established.

Therefore, the existing initial access feedback process [30] relies on a contention procedure. Specifically, BS divides the time-frequency resources into blocks. Each block is associated with a beam, and in each block, BS uses the corresponding beam to receive signal.

UE feedback is constituted by sending a contention based preamble in the block corresponding to its selected optimal beam (or the optimal wide beam), so that BS can reliably receive the preamble using the optimal beam with high beamforming gain. Thus, BS knows the selected optimal beam by UE. In other words, UE successfully provided the beam information feedback. This information will enable BS to establish its downlink for transmission to UE.

References

[1] Y. Zeng, R. Zhang, and T. J. Lim, "Wireless communications with unmanned aerial vehicles: Opportunities and challenges," *IEEE Commun. Mag.*, vol. 54, no.5, pp. 36–42, May 2016.

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

[3] W. Roh, *et al.*, "Millimeter-wave beamforming as an enabling technology for 5G cellular communications: Theoretical feasibility and prototype results," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 106–113, Feb. 2014.

[4] S. Sun, *et al.*, "MIMO for millimeter-wave wireless communications: Beamforming, spatial multiplexing, or both?" *IEEE Commun. Mag.*, vol. 52, no. 12, pp. 110–121, Dec. 2014

[5] FCC, "Millimeter wave Propagation: Spectrum Management Implications," *Office of Eng. and Tech.*, Bulletin no. 70, July, 1997.

[6] Z. Pi and F. Khan, "An introduction to millimeter-wave mobile broadband systems," *IEEE Commun. Mag.*, vol. 49, no. 6, pp. 101–107, Jun. 2011.

[7] R. W. Heath, *et al.*, "An overview of signal processing techniques for millimeter wave MIMO systems," *IEEE J. Sel. Topics Sig. Process.*, vol. 10, no. 3, pp. 436–453, Apr. 2016.

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

[9] D. De, *et al.*, "Millimeter-wave beam training acceleration through low-complexity hybrid transceivers," in *IEEE Trans. Wireless Commun.*, vol. 16, no. 6, pp. 3646–3660, Jun. 2017.

[10] 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.

[11] 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.

[12] H. Li, W. Zhang, W. Cheng, and R. Liang, "Hierarchical multi-beam search based channel estimation for millimeter-wave massive MIMO systems" *IEEE Access*, vol. 7, pp. 180684–180699. Dec. 2019.

[13] Z. Xiao, P. Xia, and X. G. Xia, "Codebook design for millimeter-wave channel estimation with hybrid precoding structure," *IEEE Trans Wireless Commun.*, vol. 16, no. 1, pp. 141–153, Jan. 2017.

[14] X. Luo, W. Liu, and Z. Wang, "Calibrated beam training for millimeter-wave massive MIMO systems," in *Proc. VTC2019-Fall* (Honolulu, HI, USA), pp. 1–5, Sep. 22-25, 2019.

[15] H. Yu, P. Guan, W. Qu and Y. Zhao, "An improved beam training scheme under hierarchical codebook," *IEEE Access*, vol. 8, pp. 53627–53635, Mar. 2020.

[16] M. Alrabeiah and A. Alkhateeb, "Deep learning for mmWave beam and blockage prediction using sub-6 GHz channels," *IEEE Trans. Commun.*, vol. 68, no. 9, pp. 5504–5518, Sep. 2020.

[17] C. Qi, Y. Wang, and G. Y Li, "Deep learning for beam training in millimeter wave massive MIMO systems," *IEEE Trans. Wireless Commun.*, early access, doi:10.1109/TWC.2020.3024279.

[18] F. B. Mismar, B. L. Evans, and A. Alkhateeb, "Deep reinforcement learning for 5G networks: Joint beamforming, power control, and interference coordination," *IEEE Trans. Commun.*, vol. 68, no. 3, pp. 1581–1592, Mar. 2020.

[19] M. Giordani, M. Mezzavilla, and M. Zorzi, "Initial access in 5G mmWave cellular networks," *IEEE Commun. Mag.*, vol. 54, no. 11, pp. 40–47, Nov. 2016.

[20] M. Giordani, *et al.*, "A tutorial on beam management for 3GPP NR at mmWave frequencies," *IEEE Commun. Surveys and Tutorials*, vol. 21, no. 1, pp. 173–196, Firstquarter 2019.

[21] R. Zhang, H. Zhang, W. Xu, and X. You, "Subarray-based simultaneous beam training for multiuser mmWave massive MIMO systems," *IEEE Wireless Commun. Lett.*, vol. 8, no. 4, pp. 976–979, Aug. 2019.

[22] C. Qi, K. Chen, O. A. Dobre, and G. Y. Li, "Hierarchical codebook-based multiuser beam training for millimeter wave massive MIMO," *IEEE Trans. Wireless Commun.*, vol. 19, no. 12, pp. 8142–8152, Dec. 2020,

[23] S. Chiu, N. Ronquillo, and T. Javidi, "Active learning and CSI acquisition for mmWave initial alignment," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 11, pp. 2474-2489, Nov. 2019.

[24] M. Hussain and N. Michelusi, "Energy-efficient interactive beam alignment for millimeter-wave networks," *IEEE Trans. Wireless Commun.*, vol. 18, no. 2, pp. 838-851, Feb. 2019.

[25] O. E. Ayach, *et al.*, "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1499–1513, Mar. 2014.

[26] P. Kumari, M. E. Eltayeb and R. W. Heath, "Sparsity-aware adaptive beamforming design for IEEE 802.11ad-based joint communication-radar," in *Proc. RadarConf18* (Oklahoma City, OK, USA), Apr. 23-27, 2018, pp. 0923–0928.

[27] 3GPP TS 38.901, "Study on channel model for frequencies from 0.5 to 100 GHz", Mar. 2022.

[28] C. T. Neil, *et al.*, "Impact of microwave and mmWave channel models on 5G systems performance," *IEEE Trans. Antennas and Propagation*, vol. 65, no. 12, pp. 6505–6520, Dec. 2017.

[29] F. Rusek, *et al.*, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Sig. Process. Mag.*, vol. 30, no. 1, pp. 40–60, Jan. 2013.

[30] *IEEE Standard for Information Technology – Telecommunications and Information Exchange Between Systems – Local and Metropolitan Area Networks – Specific Requirements – Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications Amendment 3: Enhancements for Very High Throughput in the 60 GHz Band*, IEEE Standard 802.11ad-2012, pp. 1–628, Dec. 2012.

[31] F. Dai and J. Wu, "Efficient broadcasting in ad hoc wireless networks using directional antennas," *IEEE Trans. Parallel Distrib. Syst.*, vol. 17, no. 4, pp. 335-347, Apr. 2006.

[32] A. Barg and D. Y. Nogin, "Bounds on packings of spheres in the Grassmann manifold," *IEEE Trans. Inform. Theory*, vol. 48, no. 9, pp. 2450–2454, Sep. 2002.

[33] J. H. Conway, R. H. Hardin, and N. J. A. Sloane, "Packing lines, planes, etc.: Packings in Grassmannian spaces," *Exper. Math.*, vol. 5, no. 2, pp. 139–159, 1996.

[34] D. J. Love, R. W. Heath, and T. Strohmer, "Grassmannian beamforming for multiple-input multiple-output wireless systems," *IEEE Trans. Inform. Theory*, vol. 49, no. 10, pp. 2735–2747, Oct. 2003.

[35] D. J. Love and R. W. Heath, "Equal gain transmission in multiple-input multiple-output wireless systems," *IEEE Trans. Commun.*, vol. 51, no. 7, pp. 1102–1110, Jul. 2003.

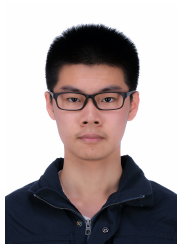
[36] D. J. Love and R. W. Heath, "Equal gain transmission in multiple-input multiple-output wireless systems," in *Proc. GLOBECOM 2002* (Taipei, Taiwan, China), Nov. 17-21, 2002, pp. 1124–1128.

[37] R. W. Heath and A. Paulraj, "A simple scheme for transmit diversity using partial channel feedback," in *Proc. 32nd Annu. Asilomar Conf. Sig., Systems, and Computers* (Pacific Grove, CA, USA), Nov. 1-4, 1998 pp. 1073–1078.

[38] B. M. Hochwald, *et al.*, "Systematic design of unitary space-time constellations," *IEEE Trans. Inform. Theory*, vol. 46, no. 6, pp. 1962–1973, Sep. 2000.

[39] M. Matthaiou, M. R. McKay, P. J. Smith, and J. A. Nossek, "On the condition number distribution of complex Wishart matrices," *IEEE Trans. Commun.*, vol. 58, no. 6, pp. 1705–1717, Jun. 2010.

Yinxiao Zhuo received the B.S. degree from the Department of Electronic Engineering, Tsinghua University, Beijing, China, in 2020, where he is currently pursuing the Ph.D. degree. His main research interest is millimeter wave communication and joint communication and sensing system.



Ziyuan Sha received the B.S. degree from the Department of Electronic Engineering, Tsinghua University, Beijing, China, in 2017, where he is currently pursuing the Ph.D. degree. His main research interest is millimeter wave communication.

Zhaocheng Wang (Fellow, IEEE) received the B.S., M.S., and Ph.D. degrees from Tsinghua University in 1991, 1993, and 1996, respectively.

From 1996 to 1997, he was a Post-Doctoral Fellow with Nanyang Technological University, Singapore. From 1997 to 1999, he was a Research Engineer/a Senior Engineer with OKI Techno Centre (Singapore) Pte. Ltd., Singapore. From 1999 to 2009, he was a Senior Engineer/a Principal Engineer with Sony Deutschland GmbH, Germany. Since 2009, he has been a Professor with the Department of Electronic Engineering, Tsinghua University, where he is currently the Director of the Broadband Communication Key Laboratory, Beijing National Research Center for Information Science and Technology (BNRist). He has authored or coauthored two books, which have been selected by IEEE Press Series on Digital and Mobile Communication (Wiley-IEEE Press). He has also authored/coauthored more than 160 peer-reviewed journal articles. He holds 46 U.S./EU granted patents (23 of them as the first inventor). His research interests include wireless communications, millimeter wave communications, and optical wireless communications.

Prof. Wang is a fellow of the Institution of Engineering and Technology. He was a recipient of the ICC2013 Best Paper Award, the OECC2015 Best Student Paper Award, the 2016 IEEE Scott Helt Memorial Award, the 2016 IET Premium Award, the 2016 National Award for Science and Technology Progress (First Prize), the ICC2017 Best Paper Award, the 2018 IEEE ComSoc AsiaPacific Outstanding Paper Award, and the 2020 IEEE ComSoc Leonard G. Abraham Prize. He was an Associate Editor of IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS from 2011 to 2015 and IEEE COMMUNICATIONS LETTERS from 2013 to 2016. He is currently an Associate Editor of IEEE TRANSACTIONS ON COMMUNICATIONS, IEEE SYSTEMS JOURNAL, and IEEE OPEN JOURNAL OF VEHICULAR TECHNOLOGY.

Sheng Chen (Fellow, IEEE) received the B.Eng. degree from East China Petroleum Institute, Dongying, China, in 1982, the Ph.D. degree from City University, London, in 1986, both in control engineering, and the Doctor of Sciences (D.Sc., the higher doctoral degree) from the University of Southampton, Southampton, U.K., in 2005.

From 1986 to 1999, he held a research and academic appointments at the Universities of Sheffield, Edinburgh, and Portsmouth, all in U.K. Since 1999, he has been with the School of Electronics and Computer Science, University of Southampton, where he holds the post of a Professor in intelligent systems and signal processing. He has more than 16,300 Web of Science citations with H-index 56 and more than 32,800 Google Scholar citations with H-index 77. His research interests include neural network and machine learning, wireless communications, and adaptive signal processing.

Dr. Chen is a fellow of the Royal Academy of Engineering, U.K., AsiaPacific Artificial Intelligence Association, and IET. He is an original ISI Highly Cited Researcher in engineering (March 2004).