Computation-Aided Adaptive Codebook Design for Millimeter Wave Massive MIMO

Yuhan Sun*, Chenhao Qi* and Geoffrey Ye Li†
*School of Information Science and Engineering, Southeast University, Nanjing, China
†School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA, USA
Email: qch@seu.edu.cn

GLOBECOM 2020 - 2020 IEEE Global Communications Conference | 978-1-7281-8298-8/20/$31.00 ©2020 IEEE | DOI: 10.1109/GLOBECOM42002.2020.9348269

***Abstract*—Different from the existing predefined hierarchical codebook before the beam training, we design a computation-aided adaptive codebook and propose a beam training algorithm based on it. At each layer of the hierarchical codebook, we first estimate the channel angle of arrival (AOA) or angle of departure (AOD) according to the beam training results from the previous layers and then adaptively design a codeword in the current layer to align with the estimated AOA or AOD. Benefiting from the computation resources used for the adaptive codebook design and beam alignment, the proposed algorithm can improve the success rate of beam training as well as reducing the training overhead comparing with the existing algorithms. Simulation results verify the effectiveness of the proposed algorithm.**

***Index Terms*—Millimeter wave (mmWave) communications, analog beamforming, hierarchical codebook, massive MIMO.**

I. Introduction

As a promising technology of next-generation wireless systems, millimeter wave (mmWave) communication has abundant spectrum resource [1]. To compensate for the severe path loss of high frequency transmission, large-scale antenna arrays use beamforming technology to concentrate beam energy in a specific direction [2].

Typically, the transmitter in mmWave massive MIMO systems employs hybrid beamformer including an analog beamformer and a digital beamformer while the receiver employs hybrid combiner including an analog combiner and a digital combiner correspondingly [3]. Analog beamforming and combining can perform directional transmission of mmWave signal by changing the phases of phase shifters connected to the antenna arrays. Digital beamformer and combiner are mainly used to tackle the interference of simultaneous transmission of multiple data streams dedicated to multiusers. Although narrower directional beams can achieve higher beam gain, it requires more accurate channel state information (CSI), e.g., angle of departure (AOD) and angle of arrival (AOA) of a mmWave channel. To obtain CSI, there are two mainstream channel training methods, including beamspace channel estimation [4] and analog beam training [5]. Beamspace channel estimation explores the sparse property of mmWave channels and treats it as a sparse recovery problem while analog beam training searches the AOA and AOD of mmWave channels by transmitting and receiving beams in different directions. An intuitive beam training method is by exhaustive search, which tests all possible transmitting and receiving beam pairs to find the best one. To improve the efficiency of exhaustive search, beam training methods based on hierarchical codebooks have been proposed. The hierarchical codebook is with multiple layers of codewords generally includes a small number of low-resolution codewords covering wide angle at upper layer of the codebook and a large number of high-resolution codewords offering high directional beamforming gain at lower layer of the codebook. In [6], a codeword design method, named SPARSE, has been proposed for a system with a large number of radio frequency (RF) chains. To reduce the number of RF chains, a hierarchical codebook design method, named JOINT, has been developed, which only needs one RF chain [7].

In this paper, we design an adaptive codebook and an efficient beam training algorithm that needs the real-time channel and training data. Different from the existing hierarchical codebook training algorithms, we first use the beam training results from the previous layers to estimate the channel AOA or AOD at each layer and then we adaptively design a codeword in the current layer to align with the estimated AOA or AOD. Comparing with the existing beam training algorithms, the proposed one can improve the success rate of beam training as well as reducing the training overhead by using the computation resources for the adaptive codebook design and beam alignment.

The notations are defined as follows. Symbols for matrices (upper case) and vectors (lower case) are in boldface. $\circ$, $(\cdot)^T$, $(\cdot)^H$, $\lfloor\cdot\rfloor$, $\lfloor\cdot\rceil$, $\mathrm{diag}\{\cdot\}$, $\boldsymbol{I}_N$, $\mathrm{E}\{\cdot\}$, $\|x\|_2$ and $\mathcal{CN}$ denote the entry-wise product, transpose, conjugate transpose (Hermitian), flooring integer operation, rounding integer operation, diagonal matrix, identity matrix of size N, expectation, ℓ_2 norm of vector x and the complex Gaussian distribution, respectively. $\mathbb{C}$ denotes the set of complex numbers.

II. System Model

After introduce the system setup in this section, we provide an overview on prior work on hierarchical codebook design.

A. System Setup

We consider an uplink mmWave massive MIMO system including a transmitter with N_t antennas and a receiver with N_r antennas. Uniform linear arrays (ULAs) with half-wavelength interval are employed for antennas at both transmitter and receiver. For single data stream transmission, the system can be modeled as

$$y = \sqrt{P}\mathbf{f}_r^H \mathbf{H}\mathbf{f}_t s + \mathbf{f}_r^H \mathbf{n}, \tag{1}$$

where $\mathbf{f}_r \in \mathbb{C}^{N_r}$ and $\mathbf{f}_t \in \mathbb{C}^{N_t}$ are the normalized analog combiner and the normalized analog beamformer, respectively, i.e., $\|\mathbf{f}_r\|_2 = 1$ and $\|\mathbf{f}_t\|_2 = 1$, P and $\mathbf{n} \in \mathbb{C}^{N_r}$ denote the total transmit power and the additive white Gaussian noise vector with $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}, \sigma^2 \boldsymbol{I}_{N_r})$, respectively. s and y denote the input symbol of the analog beamformer and the output symbol of the analog combiner, respectively. The mmWave massive MIMO channel matrix can be written as [6]

$$\mathbf{H} = \sqrt{\frac{N_t N_r}{L}} \sum_{l=1}^{L} \alpha_l \mathbf{u}(N_r, \theta_l)\mathbf{u}(N_t, \psi_l)^H \tag{2}$$

where L is the number of channel multipath, $\alpha_l \sim \mathcal{CN}(0, \sigma_l^2)$ is the channel gain of the lth path, ψ_l and θ_l denote the AOD and AOA of the lth path, respectively. The channel steering vector $\mathbf{u}(\cdot)$ can be expressed as

$$\mathbf{u}(N, \varpi) \triangleq \frac{1}{\sqrt{N}}[1, e^{j\pi\varpi}, \ldots, e^{j\pi(N-1)\varpi}]^T, \tag{3}$$

where N is the number of antennas and ϖ is the channel AOA or AOD. If we denote the physical channel AOD and physical channel AOA of the lth path as ϕ_l^t and ϕ_l^r, then we have $\psi_l = \frac{2d}{\lambda}\sin\phi_l^t$ and $\theta_l = \frac{2d}{\lambda}\sin\phi_l^r$ where λ is the wavelength of mmWave signal and we normally set the antenna interval $d \triangleq \frac{\lambda}{2}$. Both ψ_l and θ_l obey the uniform distribution between [-1,1] [8].

In fact, we have

$$\mathbf{H} = \sqrt{\frac{N_t N_r}{L}} \mathbf{U}_r \mathbf{\Lambda}_\alpha \mathbf{U}_t^H \tag{4}$$

where

$$\begin{aligned} &\mathbf{U}_r \triangleq [\mathbf{u}(N_r, \theta_1), \mathbf{u}(N_r, \theta_2), \ldots, \mathbf{u}(N_r, \theta_L)] \in \mathbb{C}^{N_r \times L}, \\ &\mathbf{U}_t \triangleq [\mathbf{u}(N_t, \psi_1), \mathbf{u}(N_t, \psi_2), \ldots, \mathbf{u}(N_t, \psi_L)] \in \mathbb{C}^{N_t \times L}, \\ &\mathbf{\Lambda}_\alpha \triangleq \mathrm{diag}\{\alpha_1, \alpha_2, \ldots, \alpha_L\} \in \mathbb{C}^{L \times L}. \end{aligned} \tag{5}$$

B. Review of Prior Work: Hierarchical Codebook Design

In mmWave massive MIMO channel, there is usually one dominant line-of-sight (LOS) path and several secondary non-line-of-sight (NLOS) paths. Hierarchical beam training usually focuses on the efficient search of the dominant path by designing hierarchical codebook layer by layer.

A hierarchical codebook $\mathbf{W} \in \mathbb{C}^{N_t \times N_w}$ with $N_w \triangleq \lceil 2(N_t - 1) \rceil$ typically covers the entire angle domain $\chi = [-1, 1]$, where different codewords from the same layer of $\mathbf{W}$ have the same resolution but different target angles. The kth codeword in the mth layer of $\mathbf{W}$ is denoted as $\mathbf{w}(m, k) \in \mathbb{C}^{N_t}$, for $m = 1, 2, \ldots, S_W$ and $k = 1, 2, \ldots, 2^m$, where $S_W \triangleq \lfloor \log_2 N_t \rfloor$ is the number of layers of $\mathbf{W}$. If we denote the main lobe of $\mathbf{w}(m, k)$ as $\beta(m, k)$, then $\beta(m, i) \cap \beta(m, k) = \emptyset, i \neq k$, which means different codewords at the same layer have different main lobes to ensure unique identification of the codewords during the beam training. We also have $\cup_{k=1}^{2^m} \beta(m, k) = \chi$, which means the main lobes of all codewords at the same layer covers the entire angle domain. The main lobe of $\mathbf{w}(m, k)$ is divided into the main lobes of two codewords in the next layer, i.e., $\beta(m, k) = \beta(m+1, 2k-1) \cup \beta(m+1, 2k)$.

Ideally, the design of codewords in a hierarchical codebook needs to meet [6],

$$|\mathbf{u}(N_t, \psi_l)^H \mathbf{w}(m, k)|^2 = \begin{cases} C_m & \psi_l \in \beta(m, k) \\ 0 & \psi_l \notin \beta(m, k) \end{cases} \tag{6}$$

where $\beta(m, k) = (-1 + \frac{2k-2}{2^m}, -1 + \frac{2k}{2^m})$. The equation (6) means that the directional gain of the same layer codewords in their the main lobe is constant and the directional gain tends to be zero if the channel path is out of the main lobe of the codewords.

In order to search the dominant path of the channel, two hierarchical codebooks, $\mathbf{W}$ and $\mathbf{V}$, need to be established at the transmitter and receiver, respectively. The beam training based on hierarchical codebooks aims to find out a pair of codewords that best matches the dominant path, which starts the search from the upper layer to the bottom layer of the hierarchical codebooks. The ith codeword in the mth layer of $\mathbf{V}$ is denoted as $\mathbf{v}(m, i)$, for $m = 1, 2, \ldots, S_V$ and $k = 1, 2, \ldots, 2^m$, where $S_V \triangleq \lfloor \log_2 N_r \rfloor$ is the number of layers of $\mathbf{V}$. Note that the small directional gain of the first layer codewords in hierarchical codebooks may bring errors to beam training. We usually start the beam training from the m_Fth layer of $\mathbf{W}$ and $\mathbf{V}$, where $m_F \geq 2$. Since there is no initial information of the channel, we need to perform an exhaustive search of codeword pairs in the m_Fth layer. After finding the codeword pair $\mathbf{w}(m, k)$ and $\mathbf{v}(m, i)$ best matching the channel in the mth layer for $m \geq m_F$, only $\mathbf{w}(m+1, 2(k-1)+p), p = 1, 2$ and $\mathbf{v}(m+1, 2(i-1)+q), q = 1, 2$ in the $(m+1)$th layer need to be tested. Therefore, the hierarchical codebook search needs to test 2^{2m_F} pairs of codewords in the m_Fth layer and 4 pairs of codewords in the other layers.

III. Computation-Aided Adaptive Codebook Design

Unlike the existing hierarchical codebook that is normally predefined before the beam training, in our work, we consider adaptive hierarchical codebook, where we first estimate the channel AOA and AOD according to the receiving signals from the previous layers of the codebooks, and then design the codewords in the current layer in an adaptive fashion. The motivation of the adaptive hierarchical codebook is to reduce the training overhead and improve the success rate of beam training by using the computation resource at the receiver, e.g., the base station in uplink transmission.

A. Computation-Aided Channel AOA and AOD Estimation

We first consider the estimation of the channel AOA, where $\mathbf{f}_t$ is fixed and the receiver using M different codewords as $\mathbf{f}_r$

for analog combining. The received signal can be expressed as

$$y_q = \sqrt{P}\mathbf{v}_q^H \mathbf{H}\mathbf{f}_t s + \mathbf{v}_q^H \mathbf{n}, \tag{7}$$

for $q = 1, 2, \ldots, M$, where $\mathbf{v}_q$ is generally defined as a codeword of $\mathbf{V}$ used in the qth signal receiving. Denote $\mathbf{y} \triangleq [y_1, y_2, \ldots, y_M]^T$ and $\mathbf{F}_r \triangleq [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_M]^H$, then we have

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{F}_r \mathbf{n}, \tag{8}$$

where $\mathbf{A} \triangleq \sqrt{N_t N_r P/L}\mathbf{F}_r \mathbf{U}_r$ and $\mathbf{x} \triangleq \mathbf{\Lambda}_\alpha \mathbf{U}_t^H \mathbf{f}_t s$. The matrix obtained by correlation processing of $\mathbf{y}$ can be expressed as

$$\mathbf{R}_\mathbf{y} \triangleq \mathbf{y}\mathbf{y}^H. \tag{9}$$

Since the rank of $\mathbf{R}_\mathbf{y}$ is one, we can get a set of eigenvectors $\mathbf{c}_i, i = 1, 2, \ldots, M$ by making the eigen decomposition of $\mathbf{R}_\mathbf{y}$, where $\mathbf{c}_1$ represents the only eigenvector with nonzero eigenvalue corresponding to the signal space. Therefore, we have

$$\mathbf{R}_\mathbf{y}\mathbf{c}_i = \mathbf{0},\ i = 2, 3, \ldots, M. \tag{10}$$

Denote the kth column of $\mathbf{A}$ as $\mathbf{a}_k$, for $k = 1, 2, \ldots, L$. We have

$$\mathbf{R}_\mathbf{y} \approx \mathbf{A}\mathbf{R}_\mathbf{x}\mathbf{A}^H = \sum_{k=1}^{L}\sum_{n=1}^{L} R_{k,n}\mathbf{a}_k\mathbf{a}_n^H \tag{11}$$

where $\mathbf{R}_\mathbf{x} \triangleq \mathbf{x}\mathbf{x}^H$ and $R_{k,n}$ denotes the entry on the kth row and nth column of $\mathbf{R}_\mathbf{x}$, for $k = 1, 2, \ldots, L$ and $n = 1, 2, \ldots, L$. Note that the approximation in (11) is based on the assumption that the noise terms are much weaker than the signal term. Ideally if there is no noise, the approximation becomes the equality. With noise, we can normally ensure that the AOA of channel LOS path lies in the main lobes of M different codewords $\mathbf{v}_q$, i.e., $\theta_1 \in \beta_q^r, q = 1, 2, \cdots, M$, where relatively higher directional gain than the noise level can be achieved to make this assumption hold. For example, we can select the best codeword covering the AOA from each layer of the hierarchical codebook and totally select M codewords from M different layers.

Based on (10) and (11), we have

$$\sum_{k=1}^{L}\sum_{n=1}^{L} R_{k,n}\mathbf{a}_k\mathbf{a}_n^H\mathbf{c}_i = \mathbf{0},\ i = 2, 3, \ldots, M. \tag{12}$$

If there is only one LOS path without any NLOS paths, i.e., $L = 1$, then (12) is simplified as

$$\mathbf{a}_1\mathbf{a}_1^H\mathbf{c}_i = \mathbf{0},\ i = 2, 3, \ldots, M. \tag{13}$$

According to the multiple signal classification (MUSIC) method, we set $\mathbf{C}_r = [\mathbf{c}_2, \mathbf{c}_3, \ldots, \mathbf{c}_M]$ and obtain an estimate of θ_1 by

$$\widehat{\theta} = \min_{\varphi\in\beta_1^r} \left\|\mathbf{F}_r\mathbf{u}(N_r, \varphi)\mathbf{u}^H(N_r, \varphi){\mathbf{F}_r}^H\mathbf{C}_r\right\|_2, \tag{14}$$

where we assume that β_1^r is the narrowest main lobe of these M different codewords. With $L > 1$, we can rewrite (12) as

$$\begin{gathered}\mathbf{a}_1\mathbf{a}_1^H\mathbf{c}_i + \tfrac{1}{R_{1,1}}\textstyle\sum_{k=1}^{L}\sum_{n=1}^{L} R_{k,n}\mathbf{a}_k\mathbf{a}_n^H\mathbf{c}_i = \mathbf{0},\\ nk \neq 1 \text{ and } i = 2, 3, \ldots, M.\end{gathered} \tag{15}$$

Since in mmWave MIMO channel, the power of the NLOS paths is much smaller than the LOS path, e.g., $\alpha_1 \sim \mathcal{CN}(0, 1)$ for the LOS path and $\alpha_2 \sim \mathcal{CN}(0, 0.01)$, $\alpha_3 \sim \mathcal{CN}(0, 0.01)$ for two NLOS paths [9], we can still use (14) to obtain an estimate of θ_1.

Similarly, we can use the method proposed above to estimate the AOD of the channel LOS path with fixed analog combiner and M different codewords $\mathbf{w}_p$ as beamformer at transmitter. The receiving signal can be expressed as

$$y_p = \sqrt{P}\mathbf{f}_r^H \mathbf{H}\mathbf{w}_p s + \mathbf{f}_r^H \mathbf{n}, \tag{16}$$

for $p = 1, 2, \ldots, M$. By defining $\mathbf{y} \triangleq [y_1, y_2, \ldots, y_M]^H$ and $\mathbf{F}_t \triangleq [\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_M]^H$, we can make eigen decomposition of $\mathbf{y}\mathbf{y}^H$ to obtain matrix $\mathbf{C}_t = [\mathbf{c}_2, \mathbf{c}_3, \ldots, \mathbf{c}_M]$. Assume that $\psi_1 \in \beta_1^t$ and β_1^t is the narrowest main lobe of these M different codewords, we can obtain an estimation of ψ_1 by

$$\widehat{\psi} = \min_{\varphi\in\beta_1^t} \left\|\mathbf{F}_t\mathbf{u}(N_t, \varphi)\mathbf{u}^H(N_t, \varphi){\mathbf{F}_t}^H\mathbf{C}_t\right\|_2. \tag{17}$$

B. Computation-Aided Adaptive Codeword Design

After estimating AOA and AOD with the aided computation, we can predict the adaptive codeword aligned with the direction of channel path in lower layer [10].

Assuming that the estimated values of the AOD and the AOA are $\widehat{\psi}$ and $\widehat{\theta}$, respectively, considering the limitation levels of a digital phase shifter, we need to quantify $\widehat{\psi}$ and $\widehat{\theta}$. We set the number of bits of digital phase shifter as D. The quantization accuracy to the angle of the steering vector is $2/2^D$, therefore the quantized values are

$$\widetilde{\psi} = \lfloor 2^{D-1}\widehat{\psi}\rceil \times 2^{1-D},\ \widetilde{\theta} = \lfloor 2^{D-1}\widehat{\theta}\rceil \times 2^{1-D}. \tag{18}$$

After the computation of the quantization angle, it is necessary to shift the codewords by different angles according to different target angles of codewords in different layers. For the S_W-layer codebook W and the S_V-layer codebook V, we can obtain the codeword adaptively by translating the first codeword in each layer, i.e., $\mathbf{w}(m, 1), m = 1, 2, \ldots, S_W$ in $\mathbf{W}$ and $\mathbf{v}(m, 1), m = 1, 2, \ldots, S_V$ in $\mathbf{V}$. According to the knowledge of codeword from Section II-B, the translation angle of $\mathbf{w}(m, 1)$ and $\mathbf{v}(m, 1)$ are $\gamma^t = \widetilde{\psi} + (2^m - 1)/2^m$ and $\gamma^r = \widetilde{\theta} + (2^m - 1)/2^m$, respectively. Using the knowledge of ULAs, the codeword can be shifted by changing the signal phase with phase shifter. Therefore, the adaptive codeword $\mathbf{w}^*(m)$ and $\mathbf{v}^*(m)$ aligned with estimated angle $\widehat{\psi}$ and $\widehat{\theta}$ in the mth layer can be expressed as

$$\begin{aligned}\mathbf{w}^*(m) &= [1, e^{j\pi\gamma^t}, \ldots, e^{j\pi(N_t-1)\gamma^t}]^T \circ \mathbf{w}(m, 1)\\ \mathbf{v}^*(m) &= [1, e^{j\pi\gamma^r}, \ldots, e^{j\pi(N_r-1)\gamma^r}]^T \circ \mathbf{v}(m, 1)\end{aligned} \tag{19}$$

where $\mathbf{w}$ and $\mathbf{v}$ are designed by the deactivation (DEACT) approach [7]. Since MUSIC method needs incoherent incident signals, we use codewords designed by DEACT approach with less correlation.

From Section II-B, it can be seen that the codewords in upper layers provides lower directional gain compared to the lower layer codewords. Therefore, it is easy to search

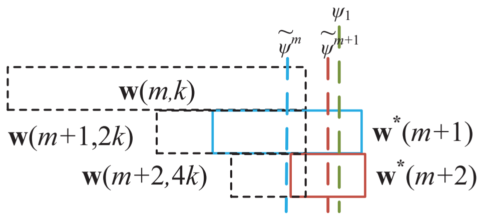

Fig. 1. Comparison of the adaptive codewords $\mathbf{w}^*$ and normal codewords $\mathbf{w}$.

incorrectly in the upper layer of the hierarchical codebook, which will lead to the failure of subsequent search. The design of adaptive codebook aims at enabling the codewords in lower layers to re-align with the dominant path after error search in upper layer by aided computation. As shown in Fig. 1, the green dotted line represents the dominant path of the channel in the direction of ψ_1, the black dotted line boxes and color boxes are the codewords in hierarchical codebook and the adaptive codewords, respectively. Assuming that when searching the mth layer, the system incorrectly regard $\mathbf{w}(m,k)$ as the best codeword, i.e., $\psi_1 \notin \beta(m,k)$. The existing hierarchical codebook will identify $\mathbf{w}(m+1,2k)$ and $\mathbf{w}(m+2,4k)$ as the best codewords in the $(m+1)$th layer and the $(m+2)$th layer, respectively, which will lead to the failure of beam training. After estimating ψ_1 and obtain $\widetilde{\psi}^m$ in the mth layer marked by the blue dotted line, the adaptive codeword $\mathbf{w}^*(m+1)$ aligned with $\widetilde{\psi}^m$ will be adopted. Then, the adaptive codeword $\mathbf{w}^*(m+2)$ generated based on the $\widetilde{\psi}^{m+1}$ estimated in the $(m+1)$th layer will be more precisely aligned with ψ_1.

C. Computation-Aided Adaptive Codebook Search Algorithm

As shown in **Algorithm 1**, we propose a computation-aided adaptive codebook design algorithm. We assume that $N_t \leq N_r$. We denote $S_W = \lfloor \log_2 N_t \rfloor$ and $S_V = \lfloor \log_2 N_r \rfloor$ as the numbers of total layers of $\mathbf{W}$ and $\mathbf{V}$, respectively. We have $S_W \leq S_V$.

1) **First stage**: We start the search of the hierarchical book from the m_Fth layer of both $\mathbf{W}$ and $\mathbf{V}$ with $2 \leq m_F \leq S_W \leq S_V$, regarding that the directional gain of the first layer codewords may not be strong enough for accurate beam training. We exhaustively test all pairs of codewords in the m_Fth layer of $\mathbf{W}$ and $\mathbf{V}$. When testing the pth codeword and the qth codeword of $\mathbf{W}$ and $\mathbf{V}$, i.e., $\mathbf{w}(m_F,p)$ and $\mathbf{v}(m_F,q)$, respectively, for $p = 1,2,\ldots,2^{m_F}$ and $q = 1,2,\ldots,2^{m_F}$, we obtain $y(p,q)$ by

$$y(p,q) = \sqrt{P}\left(\mathbf{v}(m_F,q)\right)^H \mathbf{H}\mathbf{w}(m_F,p)s + \left(\mathbf{v}(m_F,q)\right)^H \mathbf{n} \tag{20}$$

according to (1). Then we can find the best pair of p and q by

$$(p^*,q^*) = \arg\max_{p,q} |y(p,q)|. \tag{21}$$

We set $\mathbf{y} = [y(p^*,1), y(p^*,2), \ldots, y(p^*,2^{m_F})]^T$ and $\mathbf{F}_r = [\mathbf{v}(m_F,1), \mathbf{v}(m_F,2), \ldots, \mathbf{v}(m_F,2^{m_F})]^H$ and substitute them into (14) and (18) and estimate AOA as $\widetilde{\theta}^{m_F}$. Similarly, we set $\mathbf{y} = [y(1,q^*), y(2,q^*), \ldots, y(2^{m_F},q^*)]^H$ and $\mathbf{F}_t = [\mathbf{w}(m_F,1), \mathbf{w}(m_F,2), \ldots, \mathbf{w}(m_F,2^{m_F})]^H$ and substitute them into (17) and (18) and estimate AOD as $\widetilde{\psi}^{m_F}$. Then the receiver feeds $\widetilde{\psi}^{m_F}$ back to the transmitter. In practice, the receiver can feed back the index of the codewords corresponding to $\widetilde{\psi}^{m_F}$ to reduce the feedback quantity since the bottom layer codewords are normally predefined at the transmitter and receiver before the beam training.

Note that we may directly feed back p^* from the receiver to the transmitter by skipping the estimation of AOD. However, two codewords including $\mathbf{w}(m_F+1, 2p^*-1)$ and $\mathbf{w}(m_F+1, 2p^*)$ need to be tested, which doubles the testing overhead compared to that with $\widetilde{\psi}^{m_F}$ fed back.

2) **Second stage**: Based on $\widetilde{\theta}^{m_F}$, the receiver obtains $\mathbf{v}^*(m_F+1)$ via (19). Then the receiver keeps using $\mathbf{v}^*(m_F+1)$ as the analog combiner in this stage. Based on $\widetilde{\psi}^{m_F}$, the transmitter obtains $\mathbf{w}^*(m_F)$ and $\mathbf{w}^*(m_F+1)$ via (19).

Given $\widetilde{\psi}^{m-1}$ obtained from the $(m-1)$th layer of $\mathbf{W}^*$, we can obtain $\mathbf{w}^*(m)$ via (19), for $m = m_F+1,\ldots,S_W$. Then we obtain

$$y(m) = \sqrt{P}\left(\mathbf{v}^*(m_F+1)\right)^H \mathbf{H}\mathbf{w}^*(m)s + \left(\mathbf{v}^*(m_F+1)\right)^H \mathbf{n}. \tag{22}$$

Algorithm 1 Beam Training using Computation-Aided Adaptive Codebook

Input: N_t, N_r, P, m_F, $\mathbf{W}$, $\mathbf{V}$
Initialization: $S_W = \lfloor \log_2 N_t \rfloor$, $S_V = \lfloor \log_2 N_r \rfloor$.
(First Stage:)
Obtain $\widetilde{\theta}^{m_F}$ via (21), (14) and (18).
Obtain $\widetilde{\psi}^{m_F}$ via (21), (17) and (18).
Receiver feeds $\widetilde{\psi}^{m_F}$ back to Transmitter.
(Second Stage:)
Receiver obtains $\mathbf{v}^*(m_F+1)$ via (19).
for $m = m_F+1 : S_W$ **do**
  Obtain $\mathbf{w}^*(m)$ based on $\widetilde{\psi}^{m-1}$ via (19).
  Obtain a new $\widetilde{\psi}^m$ via (22), (17) and (18).
  Receiver feeds the new $\widetilde{\psi}^m$ back to Transmitter.
  if (23) is satisfied **then**
    Break.
  end if
end for
Transmitter obtains $\mathbf{w}^*(S_W)$ via (19).
(Third Stage:)
for $m = m_F+1 : S_V$ **do**
  Obtain $\mathbf{v}^*(m)$ based on $\widetilde{\theta}^{m-1}$ via (19).
  Obtain a new $\widetilde{\theta}^m$ via (24), (14) and (18).
  if (25) is satisfied **then**
    Break.
  end if
end for
Receiver obtains $\mathbf{v}^*(S_V)$ via (19).
Output: $\mathbf{f}_t = \mathbf{w}^*(S_W), \mathbf{f}_r = \mathbf{v}^*(S_V)$.

We set $\mathbf{y} = [y(m_F), y(m_F+1), \ldots, y(m)]^H$ and $\mathbf{F}_t = [\mathbf{w}^*(m_F), \mathbf{w}^*(m_F+1), \ldots, \mathbf{w}^*(m)]^H$ and substitute them into (17) and (18) and obtain a new $\widetilde{\psi}^m$. Then the receiver needs to feed the new $\widetilde{\psi}^m$ back to the transmitter after each iteration. We repeat the above procedures until the stop conditions are satisfied, which are

$$|\widehat{\psi}^m - \widehat{\psi}^{m-1}| \leq \epsilon/N_t \text{ and } |\widehat{\psi}^m - \widehat{\psi}^{m-2}| \leq \epsilon/N_t \qquad (23)$$

where $\epsilon(0 \leq \epsilon < 1)$ is a predefined convergence factor to control the convergence speed.

Once the stop conditions are satisfied based on $\widetilde{\psi}^m$ obtained in the current layer, the transmitter obtains the adaptive codeword $\mathbf{w}^*(S_W)$ via (19). Otherwise, the transmitter gets $\mathbf{w}^*(S_W)$ based on $\widetilde{\psi}^{S_W}$ obtained in the S_Wth layer via (19). Also, the receiver requires additional feedback to inform the transmitter to start the third stage.

2) **Third stage**: The transmitter keeps using $\mathbf{w}^*(S_W)$ as the analog beamformer in this stage. Based on $\widetilde{\theta}^{m_F}$, the receiver obtains $\mathbf{v}^*(m_F)$ and $\mathbf{v}^*(m_F+1)$ via (19).

Given $\widetilde{\theta}^{m-1}$ obtained from the $(m-1)$th layer of $\mathbf{V}^*$, we can obtain $\mathbf{v}^*(m)$ via (19), for $m = m_F+1, \ldots, S_V$. Then we get

$$y(m) = \sqrt{P}\,(\mathbf{v}^*(m))^H \mathbf{H}\mathbf{w}^*(S_W)s + (\mathbf{v}^*(m))^H \mathbf{n}. \qquad (24)$$

We substitute $\mathbf{y} = [y(m_F), y(m_F+1), \ldots, y(m)]^T$ and $\mathbf{F}_r = [\mathbf{v}^*(m_F), \mathbf{v}^*(m_F+1), \ldots, \mathbf{v}^*(m)]^H$ into (14) and (18) and obtain a new $\widetilde{\theta}^m$. We repeat the above procedures until the stop conditions are satisfied, where the stop conditions are

$$|\widehat{\theta}^m - \widehat{\theta}^{m-1}| \leq \epsilon/N_r \text{ and } |\widehat{\theta}^m - \widehat{\theta}^{m-2}| \leq \epsilon/N_r \qquad (25)$$

with a predefined convergence factor $\epsilon(0 \leq \epsilon < 1)$ to control the convergence speed.

Once the stop conditions are satisfied, based on $\widetilde{\theta}^m$ obtained in the current layer, the receiver obtains the adaptive codeword $\mathbf{v}^*(S_V)$ via (19). Otherwise, the receiver obtains $\mathbf{v}^*(S_V)$ based on $\widetilde{\theta}^{S_V}$ in the S_Vth layer via (19). Then the receiver feeds the signal back to the transmitter to start data transmission.

D. Overhead

Since testing a pair of codewords requires to transmit one symbol, we define the training overhead as the total number of symbols to be transmitted for beam training. Here, we adopt the hierarchical codebook in Section III-C to compare the training overhead of hierarchical codebook search algorithm and **Algorithm 1**.

According to Section II-B, the hierarchical codebook search algorithm needs to test 2^{2m_F} pairs of codewords in the m_Fth layer and 4 pairs of codewords in the other layers. When $S_V > S_W$, the hierarchical codebook search algorithm searches $\mathbf{V}$ independently, after finishing the S_Wth layer search of $\mathbf{V}$ and $\mathbf{W}$. Therefore, the training overhead of the hierarchical codebook search algorithm can be expressed as

$$T_{hc} = 2^{2m_F} + 2(S_W + S_V) - 4m_F. \qquad (26)$$

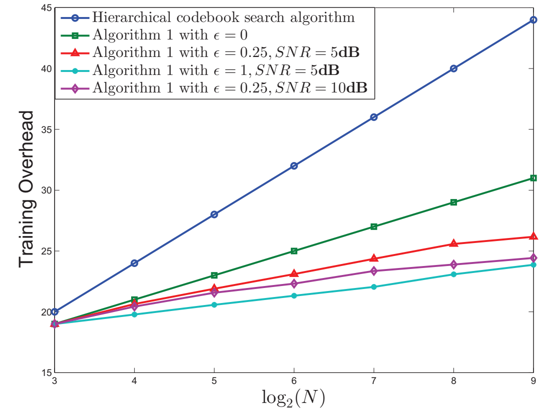

Fig. 2. Comparisons of proposed algorithm and hierarchical codebook search algorithm for different N in terms of training overhead, where $N = N_t = N_r = 2^n, n = 3, 4, \ldots, 9$ and $S_W = S_V = \log_2(N)$.

According to Section III-C, the proposed algorithm only needs to test one codeword in each layer when searching the adaptive codebook $\mathbf{W}$ and $\mathbf{V}$ separately. Hence, the training overhead of the proposed algorithm with complete search can be expressed as

$$T_{ad} = 2^{2m_F} + S_W + S_V - 2m_F. \qquad (27)$$

IV. Simulation Results

We consider an mmWave channel with $L = 3$ multipath and the AOD and AOA obey the uniform distribution $[-1, 1]$. The channel gain of the dominant path is set to be $\alpha_1 \sim \mathcal{CN}(0, 1)$, while the other two are set to be $\alpha_2 \sim \mathcal{CN}(0, 0.01)$ and $\alpha_3 \sim \mathcal{CN}(0, 0.01)$ [9].

As shown in Fig. 2, we compare the training overhead of **Algorithm 1** and the hierarchical codebook search algorithm with different antenna numbers. It can be seen that proposed algorithm outperforms the hierarchical codebook search algorithm in terms of training overhead, especially when the number of antennas is large. With the increasing number of antennas, the gap of the training overhead between the proposed algorithm and the hierarchical codebook search algorithm gets larger, e.g., the training overhead of the former can be almost half that of the latter when $N_t = N_r = 512$, $\epsilon = 1$ and SNR$=$ 5 dB. The reason is that the proposed algorithm reduces the number of transmit symbols by using aided computation to predict the adaptive codewords in the lower layers. The training overhead is further reduced when the stop conditions are used to terminate the beam training beforehand. Comparing the proposed algorithms, the training overhead decreases with the relaxation of the convergence factor ϵ and the increasing SNR. The reason is that the accuracy of aided computation estimation will increase with the SNR, which accelerates the convergence of the estimation.

Next, we compare the performance of **Algorithm 1** using the adaptive codebook and the hierarchical codebook search algorithms using JOINT codebook designed in [7] for different SNR and different ϵ. Note that JOINT codebook is the

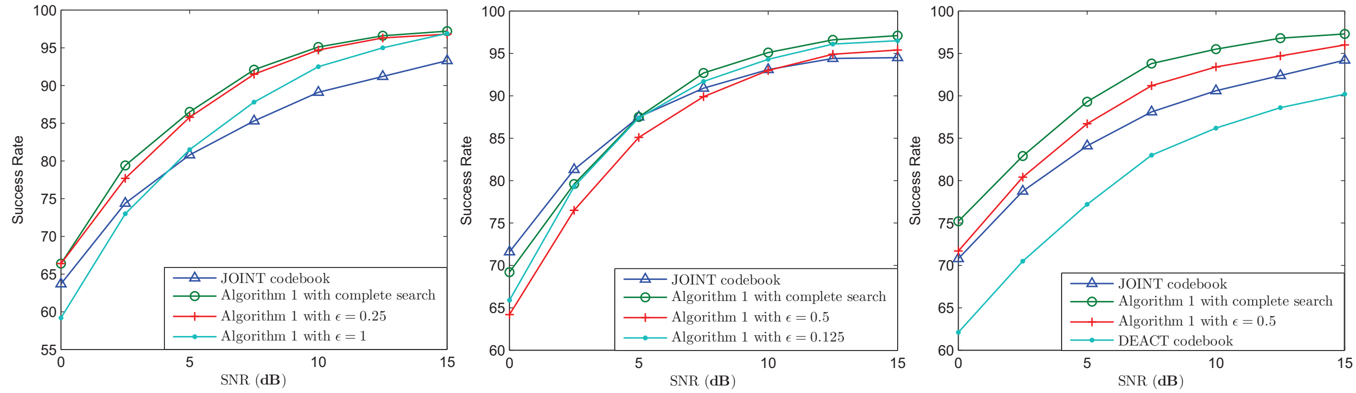

Fig. 3. Comparisons of success rate with varying SNR and ϵ. **Left** : $N_t = N_r = 32, S_W = S_V = 5$. **Middle** : $N_t = N_r = 128, S_W = S_V = 7$. **Right** : $N_t = 8, N_r = 128, S_W = 3, S_V = 7$.

state-of-the-art single RF chain codebook in terms of search success rate.

The left figure in Fig. 3 shows the success rates of **Algorithm 1** and JOINT codebook, where $N_t = N_r = 32$. From this figure, our proposed algorithm with complete search and $\epsilon = 0.25$ outperforms JOINT codebook when SNR > 0 dB, and the performance is improved by 6.8% using the proposed algorithm with complete search when SNR $= 7.5$ dB. Even if the convergence factor is relaxed, i.e., $\epsilon = 1$, the proposed algorithm still has certain advantages in performance.

The middle figure in Fig. 3 compares the success rates when $N_t = N_r = 128$. From the comparison, for SNR > 5 dB, the performance of **Algorithm 1** ($\epsilon < 0.125$) outperforms JOINT codebook and a stricter convergence factor is needed to ensure the search success rate when the number of antennas is large. As can be seen in conjunction with the left figure and middle figure in Fig. 3, JOINT codebook can achieve high success rates only when N_t, N_r are large while proposed algorithm can have high success rates in both large N_t, N_r and small N_t, N_r.

We compare the performance of codebook designed by the JOINT approach, the DEACT approach [7] and **Algorithm 1** in the right figure in Fig. 3, where $N_t = 8, N_r = 128$. It is illustrated that for a typical base station-to-user uplink mmWave massive MIMO system beam training scenario where $N_r \gg N_t$, the proposed algorithm outperforms the JOINT codebook and DEACT codebook. The gap of success rate between the proposed algorithm and the JOINT approach is up to 5.7% when SNR$= 7.5$ dB.

V. Conclusions

In this paper, we have proposed an algorithm to design adaptive codebook using computation resources to reduce the training overhead and improve the training performance. Simulation results show that the proposed algorithm outperforms the hierarchical codebook search algorithm, where the training overhead of the former is almost half that of the latter when the number of antennas at the transmitter and the receiver is equal to 512. Moreover, compared with the state-of-the-art single RF chain codebook design, the proposed algorithm has a higher search success rate under the SNR commonly used in beam training.

Acknowledgment

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

References

[1] B.-L. Wang, F.-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, 2018.

[2] J. A. Zhang, X. Huang, V. Dyadyuk, and Y. J. Guo, "Massive hybrid antenna array for millimeter-wave cellular communications," *IEEE Trans. Wireless Commun.*, vol. 22, no. 1, pp. 79–87, Mar. 2015.

[3] R. W. Heath, N. Gonzlez-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.

[4] 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.

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

[6] A. Alkhateeb, O. E. Ayach, G. Leus, and R. W. Heath, "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.

[7] 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.

[8] W. Shen, L. Dai, Y. Li, Z. Wang, and L. Hanzo, "Channel feedback codebook design for millimeter-wave massive mimo systems relying on lens antenna array," *IEEE Wireless Commun. Lett.*, vol. 7, no. 5, pp. 736–739, Oct. 2018.

[9] X. Gao, L. Dai, Z. Chen, Z. Wang, and Z. Zhang, "Near-optimal beam selection for beamspace mmwave massive MIMO systems," *IEEE Commun. Lett.*, vol. 20, no. 5, pp. 1054–1057, May 2016.

[10] D. Zhu, J. Choi, and R. W. Heath, "Auxiliary beam pair enabled aod and aoa estimation in closed-loop large-scale millimeter-wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 16, no. 7, pp. 4770–4785, May 2017.