Correlation Based Adaptive Compressed Sensing for Millimeter Wave Channel Estimation

Jianyi Yang[1,2], Zaixue Wei[1], Xin Zhang[1], Nanxi Li[1] and Lin Sang[1]
[1]School of Information and Communication Engineering, Beijing University of Posts and Telecommunications
[2]Beijing Advanced Innovation Center for Future Internet Technology
Email: {yangjianyi, zaixuew, zhangxin, airnanxil, sanglin}@bupt.edu.cn

***Abstract*—In millimeter wave (mmWave) systems, the low-overhead and high-accuracy estimation of channel state information (CSI), especially the angle of departure (AoD) and the angle of arrival (AoA), is essential for hybrid beamforming. The existing adaptive compressed sensing (ACS) based channel estimation algorithm decides the angles by comparing the received powers of different beams. The estimated angular resolution of this power based method is limited by the resolution of beams. In this paper, a correlation based adaptive compressed sensing (CB-ACS) method is proposed. It decides the angles by comparing the correlations between the received vector and the quantized sensing vectors. For the proposed method, two criteria for the beam pattern design are proposed and the cosine beam pattern is proved to be suitable. Then the corresponding hierarchical codebook is designed considering the hardware constraints. Simulation results show that with the proposed scheme, more precise estimation can be realized without increasing the training overhead. Thus, the achievable spectral efficiency of the hybrid precoding system increases.**

I. Introduction

Millimeter wave (mmWave) communications are very promising for the next generation wireless systems owing to its large available bandwidth [1]. To compensate for the severe path loss of mmWave channel, massive multiple-input multiple-output (MIMO) are adopted [2]. In this kind of systems, to reduce the cost and energy consumption, the hybrid analog and digital beamforming/precoding structure with limited radio frequency (RF) chains is usually employed. Moreover, the RF beamformer is implemented by finite-resolution analog phase shifters [3].

However, for mmWave massive MIMO systems, the low-overhead and high-accuracy estimation of the channel state information (CSI), which is essential for the precoder and combiner design, is a challenging problem [4]. Fortunately, leveraging the sparse feature of mmWave channel, channel estimation can be formulated as the sparse recovery problem and can be solved by compressed sensing (CS) tools with relatively low training overhead [5].

One application of CS is the adaptive estimation method in [6]-[8]. The efficient adaptive compressed sensing (ACS) based algorithm is proposed in [6]. It narrows down the range of angle of departure (AoD) and angle of arrival (AoA) by comparing the received powers of multi-resolution beams in each level. On the basis of [6], [7] improves the multi-resolution hierarchical codebook design by exploiting the sub-array and deactivation antenna processing. The enhanced codebook design requires fewer RF chains and activates more than half of the antennas. In addition, leveraging the binary hypothesis testing model, [8] designs the multilevel beamforming sequence to improve the data rate performance. For this power based comparison in [6]-[8], the estimated angle resolution is determined by the resolution of beams.

On the other hand, [9] and [10] estimate the mmWave channel via orthogonal matching pursuit (OMP) [11]. Different from the power-based comparison, OMP algorithm decides AOD/AOA by comparing the correlation between the received vector and the quantized sensing vectors. In [9], the training codes are designed by minimizing the total coherence of the sensing matrix and large coherence values are spread out by randomly permuting RF precoding vectors. Thus, desirable estimation accuracy is obtained. However, the computation complexity is huge, which is unaffordable at mobile station (MS). A multi-grid OMP algorithm proposed in [10] reduces the computation complexity though at the cost of some accuracy. Moreover, a two-stage CS algorithm has been proposed in [12] as a compromise of the adaptive method and OMP, exploiting the random analog beamforming.

In this paper, in order to raise the estimation accuracy without increasing the training overhead, we develop the correlation based adaptive compressed sensing (CB-ACS) channel estimation algorithm and analyze the training overhead and computation complexity of it. Besides, considering the directivity and correlation requirement, we give two criteria for the beam pattern and prove that cosine beam pattern is suitable, followed by the design of the hierarchical codebook. Simulation results show that without increasing the training overhead, higher estimation accuracy and higher spectral efficiency can be realized by CB-ACS.

Symbol Notations:$\boldsymbol{A}^*, \boldsymbol{A}^{\mathrm{T}}, \boldsymbol{A}^{\mathrm{H}}, \boldsymbol{A}^{-1}$ are the conjugate, the transport, the conjugate transport and the inverse of a matrix $\boldsymbol{A}$. $\otimes$ and $\circ$ denote the Kronecker product and the Hadamard product. $\mathrm{diag}\left(\boldsymbol{A}_1, \boldsymbol{A}_2, \cdots, \boldsymbol{A}_N\right)$ represents a block diagonal matrix whose diagonal entries are given by $\{\boldsymbol{A}_1, \boldsymbol{A}_2, \cdots, \boldsymbol{A}_N\}$. $\mathrm{vec}\left(\cdot\right)$ denotes a vector obtained through vectorizing a matrix $\boldsymbol{A}$ by column. $\mathrm{cmod}\left(x, M\right)$ equals the remainder of x divided by M if the remainder does not equal zero. Else, it equals M. $\mathrm{fix}\left(x\right)$ is the integer nearest to x towards zero. $\mathrm{dis}\left(x, y, M\right)$ equals $|x-y|$ if $|x-y| < M/2$. Else, it equals $M-|x-y|$. $\mathcal{CN}\left(\boldsymbol{a}, \boldsymbol{B}\right)$ denotes the complex Gaussian vector with mean $\boldsymbol{a}$ and covariance matrix $\boldsymbol{B}$.

978-1-5090-4183-1/17/$31.00 ©2017 IEEE

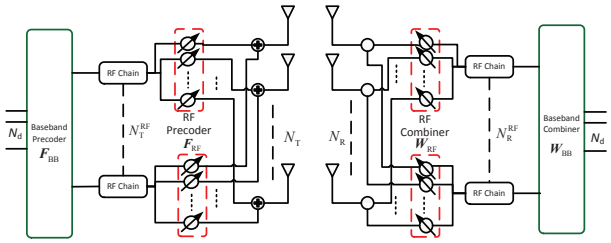

Fig. 1. Transceiver structure at BS and MS.

II. System Model and Formulation of mmWave Channel Estimation Problem

A. System Model

Consider the point to point downlink system model with large antenna arrays. The hybrid precoding/combining structure [6] is shown in Fig. 1. The base station (BS) equips N_{T} antennas and $N_{\mathrm{T}}^{\mathrm{RF}}$ RF chains while the mobile station (MS) equips N_{R} antennas and $N_{\mathrm{R}}^{\mathrm{RF}}$ RF chains. Assume that N_{d} is the number of data streams. The baseband precoding matrix is $\boldsymbol{F}_{\mathrm{BB}} \in \mathbb{C}^{N_{\mathrm{T}}^{\mathrm{RF}} \times N_{\mathrm{d}}}$, and the RF precoding matrix is $\boldsymbol{F}_{\mathrm{RF}} \in \mathbb{C}^{N_{\mathrm{T}} \times N_{\mathrm{T}}^{\mathrm{RF}}}$ whose entries satisfy $\left|[\boldsymbol{F}_{\mathrm{RF}}]_{m,n}\right|^2 = \frac{1}{N_{\mathrm{T}}}$. The precoder at BS is $\boldsymbol{F} = \boldsymbol{F}_{\mathrm{RF}}\boldsymbol{F}_{\mathrm{BB}}$ that satisfies $\|\boldsymbol{F}\|_{\mathrm{F}}^2 = N_{\mathrm{d}}$. $\boldsymbol{W} = \boldsymbol{W}_{\mathrm{RF}}\boldsymbol{W}_{\mathrm{BB}}$ is the combiner at MS which is similar as the precoder at BS.

Suppose that $\boldsymbol{s} \in \mathbb{C}^{N_{\mathrm{d}} \times 1}$ is the transmitted symbol vector such that $\mathbb{E}\left[\boldsymbol{s}\boldsymbol{s}^{\mathrm{H}}\right] = \frac{1}{N_{\mathrm{d}}}\boldsymbol{I}_{N_{\mathrm{d}}}$, where $\boldsymbol{I}_N$ denotes an $N \times N$ identity matrix. The received signal observed by MS is

$$\boldsymbol{y} = \sqrt{P_{\mathrm{d}}}\boldsymbol{W}^{\mathrm{H}}\boldsymbol{H}\boldsymbol{F}\boldsymbol{s} + \boldsymbol{W}^{\mathrm{H}}\boldsymbol{n}, \tag{1}$$

where P_{d} is the average total transmit power, $\boldsymbol{H}$ is the channel matrix and $\boldsymbol{n} \sim \mathcal{CN}\left(\boldsymbol{0}, \sigma^2\boldsymbol{I}_{N_{\mathrm{R}}}\right)$ is the Gaussian noise vector. The two dimension (2D) geometric channel model [6] is adopted in this paper. The channel matrix can be expressed as

$$\boldsymbol{H} = \sqrt{\frac{N_{\mathrm{T}}N_{\mathrm{R}}}{\rho L}}\sum_{l=1}^{L}\alpha_l\boldsymbol{a}_{\mathrm{R}}\left(\theta_{\mathrm{R},l}\right)\boldsymbol{a}_{\mathrm{T}}^{\mathrm{H}}\left(\theta_{\mathrm{T},l}\right), \tag{2}$$

where L denotes the number of scattering paths, ρ denotes the average path-loss, $\alpha_l \sim \mathcal{CN}\left(0, \sigma_\alpha^2\right)$ denotes the complex gain of the lth path, $\boldsymbol{a}_{\mathrm{R}}$ and $\boldsymbol{a}_{\mathrm{T}}$ are the array response vectors (ARVs) at BS and MS, respectively. Assuming that the adjacent antenna distance is half of the wavelength, $\boldsymbol{a}_{\mathrm{T}}\left(\theta_{\mathrm{T},l}\right)$ can be expressed as

$$\boldsymbol{a}_{\mathrm{T}}\left(\theta_{\mathrm{T},l}\right) = \frac{1}{\sqrt{N_{\mathrm{T}}}}\left[1, \mathrm{e}^{j\pi\theta_{\mathrm{T},l}}, \cdots, \mathrm{e}^{j\pi(N_{\mathrm{T}}-1)\theta_{\mathrm{T},l}}\right]^{\mathrm{T}}, \tag{3}$$

where $\theta_{\mathrm{T},l}$ is the cosine value of AoD. $\boldsymbol{a}_{\mathrm{R}}\left(\theta_{\mathrm{R},l}\right)$ can be expressed as the similar style where $\theta_{\mathrm{R},l}$ is the cosine value of AoA. For simplicity, we term θ_{T} and θ_{R} AoD and AoA respectively in the following.

The achievable spectral efficiency (ASE) of the mmWave channel can be denoted as [4]

$$R = \log_2\left(\left|\boldsymbol{I}_{N_{\mathrm{d}}} + \frac{P_{\mathrm{d}}}{N_{\mathrm{d}}}\boldsymbol{R}_n^{-1}\boldsymbol{W}^{\mathrm{H}}\boldsymbol{H}\boldsymbol{F}\boldsymbol{F}^{\mathrm{H}}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{W}\right|\right), \tag{4}$$

where $\boldsymbol{R}_n = \sigma_n^2\boldsymbol{W}^{\mathrm{H}}\boldsymbol{W}$ is the noise covariance matrix after combining. Besides, the transmitting signal to noise ratio (SNR) is defined as $\gamma_{\mathrm{d}} = \frac{P_{\mathrm{d}}}{\rho\sigma^2}$.

B. Formulation of mmWave Channel Estimation Problem

To estimate the mmWave channel, we first describe the quantized channel [9]. The quantized spatial angles, also called the grids, is defined as

$$\bar{\Theta} = \left\{\bar{\theta}_g \mid \bar{\theta}_g = -1 + \frac{2}{G}\left(g-1\right), g = 1, 2, \cdots, G\right\}, \tag{5}$$

where G denotes the number of the quantized angles and satisfies $G \gg L$. Then, the channel matrix can be represented as [9]

$$\boldsymbol{H} = \bar{\boldsymbol{A}}_{\mathrm{R}}\bar{\boldsymbol{H}}_{\mathrm{d}}\bar{\boldsymbol{A}}_{\mathrm{T}}^{\mathrm{H}} + \boldsymbol{E}, \tag{6}$$

where $\bar{\boldsymbol{A}}_{\mathrm{T}} = \left[\boldsymbol{a}_{\mathrm{T}}\left(\bar{\theta}_1\right), \boldsymbol{a}_{\mathrm{T}}\left(\bar{\theta}_2\right), \cdots, \boldsymbol{a}_{\mathrm{T}}\left(\bar{\theta}_G\right)\right]$ is the collection of quantized ARVs. $\bar{\boldsymbol{H}}_{\mathrm{d}} \in \mathbb{C}^{G \times G}$ is a sparse matrix with L non-zero entries corresponding to L paths, and $\boldsymbol{E}$ is the quantization error matrix.

When we measure the mmWave downlink channel, BS uses M_{T} training vectors and MS uses M_{R} ones. If BS uses a precoding vector $\boldsymbol{f}_p$ and MS uses a combining vector $\boldsymbol{w}_q$, the received signal can be expressed as $\boldsymbol{y}_{p,q} = \boldsymbol{w}_q^{\mathrm{H}}\boldsymbol{H}\boldsymbol{f}_p s_p + \boldsymbol{w}_q^{\mathrm{H}}\boldsymbol{n}_{p,q}$. Assuming that all the signals equal $\sqrt{P_{\mathrm{t}}}$, the received signal is

$$\boldsymbol{Y} = \sqrt{P_{\mathrm{t}}}\boldsymbol{W}^{\mathrm{H}}\boldsymbol{H}\boldsymbol{F} + \mathrm{diag}\left(\boldsymbol{w}_1^{\mathrm{H}}, \boldsymbol{w}_2^{\mathrm{H}}, \cdots, \boldsymbol{w}_{M_{\mathrm{R}}}^{\mathrm{H}}\right)\boldsymbol{N}, \tag{7}$$

where $\boldsymbol{W} = [\boldsymbol{w}_1, \boldsymbol{w}_2, \cdots, \boldsymbol{w}_{M_{\mathrm{R}}}]$ is the collection of training vectors at MS, $\boldsymbol{F} = \left[\boldsymbol{f}_1, \boldsymbol{f}_2, \cdots, \boldsymbol{f}_{M_{\mathrm{T}}}\right]$ is the collection of training vectors at BS, $\boldsymbol{N} = \Big[\left[\boldsymbol{n}_{11}^{\mathrm{T}}, \boldsymbol{n}_{12}^{\mathrm{T}}, \cdots, \boldsymbol{n}_{1M_{\mathrm{R}}}^{\mathrm{T}}\right]^{\mathrm{T}}, \left[\boldsymbol{n}_{21}^{\mathrm{T}}, \boldsymbol{n}_{22}^{\mathrm{T}}, \cdots, \boldsymbol{n}_{2M_{\mathrm{R}}}^{\mathrm{T}}\right]^{\mathrm{T}}, \cdots, \left[\boldsymbol{n}_{M_{\mathrm{T}}1}^{\mathrm{T}}, \boldsymbol{n}_{M_{\mathrm{T}}2}^{\mathrm{T}}, \cdots, \boldsymbol{n}_{M_{\mathrm{T}}M_{\mathrm{R}}}^{\mathrm{T}}\right]^{\mathrm{T}}\Big]$ is the noise matrix. The SNR when estimating the channel is $\gamma_{\mathrm{t}} = \frac{P_{\mathrm{t}}}{\rho\sigma^2}$.

To get the sparse expression of the received signal, we vectorize the received matrix $\boldsymbol{Y}$ as [6]

$$\begin{aligned}\boldsymbol{y}_{\mathrm{v}} &= \sqrt{P_{\mathrm{t}}}\mathrm{vec}\left(\boldsymbol{W}^{\mathrm{H}}\boldsymbol{H}\boldsymbol{F}\right) + \boldsymbol{n}_{\mathrm{v}} \\ &= \sqrt{P_{\mathrm{t}}}\left(\boldsymbol{F}^{\mathrm{T}}\bar{\boldsymbol{A}}_{\mathrm{T}}^{*} \otimes \boldsymbol{W}^{\mathrm{H}}\bar{\boldsymbol{A}}_{\mathrm{R}}\right)\boldsymbol{x} + \boldsymbol{e}_{\mathrm{v}} + \boldsymbol{n}_{\mathrm{v}},\end{aligned} \tag{8}$$

where $\boldsymbol{x} = \mathrm{vec}\left(\boldsymbol{H}_{\mathrm{d}}\right)$ is a vector with L non-zero elements, $\boldsymbol{e}_{\mathrm{v}}$ is the quantization error matrix, $\boldsymbol{n}_{\mathrm{v}}$ is the noise vector.

If the sensing matrix is denoted as $\boldsymbol{Q} = \boldsymbol{F}^{\mathrm{T}}\bar{\boldsymbol{A}}_{\mathrm{T}}^{*} \otimes \boldsymbol{W}^{\mathrm{H}}\bar{\boldsymbol{A}}_{\mathrm{R}}$, the channel estimation problem can be written as the famous sparse recovery problem [9]

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}}\left\|\boldsymbol{y}_{\mathrm{v}} - \sqrt{P_{\mathrm{t}}}\boldsymbol{Q}\boldsymbol{x}\right\|_2, \text{subject to } \|x\|_0 = L, \tag{9}$$

where $\|\cdot\|_0$ denotes the number of non-zero elements of a vector. $\boldsymbol{q}\left(\bar{\theta}_{g_1}, \bar{\theta}_{g_2}\right)$, called the quantized sensing vector, is one of the columns in Q and corresponding to two angles $\bar{\theta}_{g_1}$ and $\bar{\theta}_{g_2}$. According to the correlation, for a single-path channel, the AoD and AoA can be decided as [11]

$$\left(\hat{\theta}_{\mathrm{T}}, \hat{\theta}_{\mathrm{R}}\right) = \arg\max_{\left(\bar{\theta}_{g_1}, \bar{\theta}_{g_2}\right)}\left|\boldsymbol{y}_{\mathrm{v}}^{\mathrm{H}}\boldsymbol{q}\left(\bar{\theta}_{g_1}, \bar{\theta}_{g_2}\right)\right|. \tag{10}$$

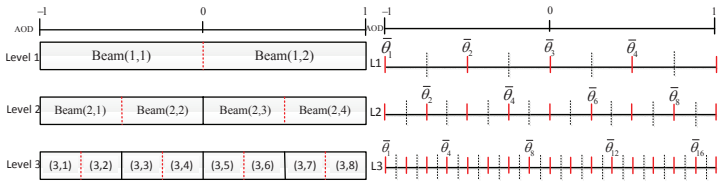

Fig. 2. Division of the angle domain by power.

Fig. 3. Division of the angle domain by correlation.

III. The Correlation Based Adaptive Compressed Sensing Algorithm

Generally, to estimate a path, ACS is performed for several levels. A predefined hierarchical training codebook is needed. Assume that in the sth $(s = 1, 2, \cdots, S)$ level, $\boldsymbol{F}_s$ is the collection of all M_s precoding vectors and $\boldsymbol{W}_s$ is the collection of all M_s combining vectors. Note that the number of levels is $S = \log_K (M_S/M_1) + 1$, where $K = M_s/M_{s-1}$. The training of the first level is also called the coarse training, where all M_1 precoding/combining vectors in the first level are used to sense the whole angle domain. The training of the following levels is also called the precise training. In the sth $(s = 2, \cdots, S)$ level, M_P precoding or combining vectors in $\boldsymbol{F}_s$ or $\boldsymbol{W}_s$ are used. M_P can be much smaller than M_s by exploiting the result of the last level. To estimate L_d paths, the training overhead is

$$T = \left(M_1^2 + (S-1)M_\mathrm{P}^2\right) L_\mathrm{d}. \tag{11}$$

In addition, $G_s (s = 1, 2, \cdots, S)$ is defined as the angular resolution of the sth level. So the final angular resolution G is G_S.

Before the CB-ACS algorithm, the previous ACS method [6], called the power based adaptive compressed sensing (PB-ACS) algorithm by us, is reviewed. Its hierarchical structure is shown in Fig. 2. In each level, the selected angle interval is divided by beams and one interval is selected according to the received power. Thus, M_1 beams are used in the coarse training stage and $M_\mathrm{P} = K$ beams are used in each level of precise training. In the sth level, The angular resolution determined by the number of beams is $G_s = M_s = M_1 K^{s-1}$.

Different from the PB-ACS, CB-ACS divide the angle domain according to the correlation between the received vector and the quantized sensing vectors. Each quantized sensing vector is corresponding to a quantized AoD and a quantized AoA. The hierarchical AoD grids are marked by the red lines in Fig. 3. The angular resolution G_s is the number of grids in this level. In the first $(S-1)$ levels, G_s equals M_{s+1} so that the grids in the sth level and the beams in the $(s+1)$th level are corresponding one by one. In the last level, G_S equals the required resolution G.

The CB-ACS algorithm is described in detail in Algorithm 1. Besides the codebook, the quantized ARVs $\bar{A}_{\mathrm{T},s}$ and $\bar{A}_{\mathrm{R},s}$ are assumed to be known by BS and MS. L_d is the number of paths to be estimated. In the coarse training step, all M_1 codes in $\boldsymbol{F}_1$ and $\boldsymbol{W}_1$ are used. After measuring, the contributions of the estimated paths are projected out and the residue of the received vector $\boldsymbol{y}_\mathrm{r}$ is obtained. Then, the index of the sensing vector, which has the largest correlation with the received vector, is selected. According to this index, the grid indices $i_{\mathrm{T},1}$ and $i_{\mathrm{R},1}$ are obtained and fed back to the BS.

Algorithm 1 Correlation based adaptive compressed sensing algorithm

Input: BS and MS have $\boldsymbol{F}_s$,$\boldsymbol{W}_s$, $\bar{\boldsymbol{A}}_{\mathrm{T},s}$,$\bar{\boldsymbol{A}}_{\mathrm{R},s}$,$L_\mathrm{d}$, M_P

for $l = 1 : L_\mathrm{d}$ **do**
  Coarse training: BS uses $\boldsymbol{F}_1$ while MS uses $\boldsymbol{W}_1$, the received vector is denoted as $\boldsymbol{y}$.
  for $p = 1 : l-1$ **do**
    $\hat{\boldsymbol{q}}_p = \boldsymbol{F}_1^\mathrm{T} \boldsymbol{a}_\mathrm{T}^* \left(\hat{\theta}_{\mathrm{T},p}\right) \otimes \boldsymbol{W}_1^\mathrm{H} \boldsymbol{a}_\mathrm{R} \left(\hat{\theta}_{\mathrm{R},p}\right)$
  end for
  $\boldsymbol{y}_\mathrm{r} = \boldsymbol{y} - \hat{\boldsymbol{Q}} \left(\hat{\boldsymbol{Q}}^\mathrm{H} \hat{\boldsymbol{Q}}\right)^{-1} \hat{\boldsymbol{Q}}^\mathrm{H} \boldsymbol{y}$, $\hat{\boldsymbol{Q}} = \left[\hat{\boldsymbol{q}}_1, \hat{\boldsymbol{q}}_2, \cdots, \hat{\boldsymbol{q}}_{l-1}\right]$
  $i_\mathrm{C} = \arg\max_i \left|[\boldsymbol{Q}_1]_{:,i}^\mathrm{H} \boldsymbol{y}_\mathrm{r}\right|$, $\boldsymbol{Q}_1 = \boldsymbol{F}_1^\mathrm{T} \bar{\boldsymbol{A}}_{\mathrm{T},1}^* \otimes \boldsymbol{W}_1^\mathrm{H} \bar{\boldsymbol{A}}_{\mathrm{R},1}$
  $i_{\mathrm{T},1} = \mathrm{fix}\left(i_\mathrm{C}/G_1\right) + 1$, $i_{\mathrm{R},1} = i_\mathrm{C} - \left(i_{\mathrm{T},1} - 1\right) G_1$
  Precise training:
  for $s = 2 : S$ **do**
    BS uses $\boldsymbol{F}_{\mathrm{P},s} = [\boldsymbol{F}_s]_{:,\mathcal{I}_{\mathrm{T},s}}$ while MS uses $\boldsymbol{W}_{\mathrm{P},s} = [\boldsymbol{W}_s]_{:,\mathcal{I}_{\mathrm{R},s}}$, the received vector is $\boldsymbol{y}$.
    for $p = 1 : l-1$ **do**
      $\hat{\boldsymbol{q}}_p = \boldsymbol{F}_{\mathrm{p},s}^\mathrm{T} \boldsymbol{a}_\mathrm{T}^* \left(\hat{\theta}_{\mathrm{T},p}\right) \otimes \boldsymbol{W}_{\mathrm{p},s}^\mathrm{H} \boldsymbol{a}_\mathrm{R} \left(\hat{\theta}_{\mathrm{R},p}\right)$
    end for
    $\boldsymbol{y}_\mathrm{r} = \boldsymbol{y} - \hat{\boldsymbol{Q}} \left(\hat{\boldsymbol{Q}}^\mathrm{H} \hat{\boldsymbol{Q}}\right)^{-1} \hat{\boldsymbol{Q}}^\mathrm{H} \boldsymbol{y}$, $\hat{\boldsymbol{Q}} = \left[\hat{\boldsymbol{q}}_1, \hat{\boldsymbol{q}}_2, \cdots, \hat{\boldsymbol{q}}_{l-1}\right]$
    $\bar{\boldsymbol{A}}_{\mathrm{PT},s} = \left[\bar{\boldsymbol{A}}_{\mathrm{T},s}\right]_{:,\mathcal{J}_{\mathrm{T},s}}$, $\bar{\boldsymbol{A}}_{\mathrm{PR},s} = \left[\bar{\boldsymbol{A}}_{\mathrm{R},s}\right]_{:,\mathcal{J}_{\mathrm{R},s}}$
    $i_\mathrm{P} = \arg\max_i \left|[\boldsymbol{Q}_s]_{:,i}^\mathrm{H} \boldsymbol{y}_\mathrm{r}\right|$, $\boldsymbol{Q}_s = \boldsymbol{F}_{\mathrm{P},s}^\mathrm{T} \bar{\boldsymbol{A}}_{\mathrm{PT},s}^* \otimes \boldsymbol{W}_{\mathrm{P},s}^\mathrm{H} \bar{\boldsymbol{A}}_{\mathrm{PR},s}$
    $i_\mathrm{PT} = \mathrm{fix}\left(i_\mathrm{P}/G_\mathrm{P}\right) + 1$, $i_\mathrm{PR} = i_\mathrm{P} - \left(i_\mathrm{PT} - 1\right) G_\mathrm{P}$
    $i_{\mathrm{T},s} = \left[\mathcal{J}_{\mathrm{T},s}\right]_{i_\mathrm{PT}}$, $i_{\mathrm{R},s} = \left[\mathcal{J}_{\mathrm{R},s}\right]_{i_\mathrm{PR}}$
  end for
  $\hat{\theta}_{\mathrm{T},l} = -1 + \frac{2}{G_S}\left(i_{\mathrm{T},S} - 1\right)$, $\hat{\theta}_{\mathrm{R},l} = -1 + \frac{2}{G_S}\left(i_{\mathrm{R},S} - 1\right)$
  $\hat{\alpha}_l = \left(\boldsymbol{q}^\mathrm{H} \boldsymbol{q}\right)^{-1} \boldsymbol{q}^\mathrm{H} \boldsymbol{y}$, $\boldsymbol{q} = \boldsymbol{F}_{\mathrm{P},S}^\mathrm{T} \boldsymbol{a}_\mathrm{T}^* \left(\hat{\theta}_{\mathrm{T},l}\right) \otimes \boldsymbol{W}_{\mathrm{P},S}^\mathrm{H} \boldsymbol{a}_\mathrm{R} \left(\hat{\theta}_{\mathrm{R},l}\right)$
end for

At the precise training step, given the required resolution, the training beam number M_P is set near K. For similarity, a grid in the $(s-1)$th level and its corresponding beam in the sth level have the same index. In this way, the transmitting beam index collection is $\mathcal{I}_{\mathrm{T},s} = \{\mathrm{cmod}\left(i_{\mathrm{T},s-1} + n, M_s\right)\}$ where $n = -\frac{M_\mathrm{P}}{2}, -\frac{M_\mathrm{P}}{2} + 1, \cdots, \frac{M_\mathrm{P}}{2} - 1$, if M_P is even, and else, $n = -\frac{M_\mathrm{P}-1}{2}, -\frac{M_\mathrm{P}-1}{2} + 1, \cdots, \frac{M_\mathrm{P}-1}{2}$. $\mathcal{I}_{\mathrm{R},s}$ is similar. Since $G_\mathrm{P} = \frac{G_s}{M_s}\left(M_\mathrm{P} - 1\right) + 1$ grids are in the coverage of the transmitting beams, the alternative collection of the grid indices is $\mathcal{J}_{\mathrm{T},s} = \left\{\mathrm{cmod}\left(j_{\mathrm{T},s} + n, G_s\right), n = -\frac{G_\mathrm{P}-1}{2}, -\frac{G_\mathrm{P}-1}{2} + 1, \cdots, \frac{G_\mathrm{P}-1}{2}\right\}$, where $j_{\mathrm{T},s} = \frac{G_s}{M_s} i_{\mathrm{T},s-1} - \frac{G_s}{M_s} + 1$. $\mathcal{J}_{\mathrm{R},s}$ is similar. $i_{\mathrm{T},s}$ and $i_{\mathrm{R},s}$ are worked out in a similar way with the coarse training and fed back to BS. After $S-1$ stages, the AOD $\hat{\theta}_{\mathrm{T},l}$ and AOA $\hat{\theta}_{\mathrm{R},l}$ of this path are worked out according to $i_{\mathrm{T},S}$ and $i_{\mathrm{R},S}$. Then the complex gain of this path $\hat{\alpha}_l$ is calculated by least squares (LS).

Since M_1 and M_P can be very small, the training over-

head of CB-ACS is much less than the exhaustive training overhead $N_{\mathrm{T}}N_{\mathrm{R}}$. To complete traditional OMP, $L_{\mathrm{d}}G_S^2$ correlation computations are needed, which is unaffordable when a precise estimation is required. But for CB-ACS, only $\left(G_1^2+(S-1)\,G_{\mathrm{P}}^2\right)L_{\mathrm{d}}$ correlation computations are needed. The computational complexity is reduced to a great extent.

On the other hand, to raise the energy efficiency, PA is one of the options [6]. Transmitting power is allocated to compensate for the low array gain in primary levels with the average power P_{t} constant. For CB-ACS, if the maximum array gain in the sth level is C_s, the transmitting power of the sth level is

$$P_s = P_{\mathrm{t}}\frac{M_1^2+(S-1)\,M_{\mathrm{P}}^2}{M_1^2C_s^2/C_1^2+C_s^2\sum\limits_{n=2}^{S}\left(M_{\mathrm{P}}^2/C_n^2\right)}. \tag{12}$$

IV. Hierarchical Codebook Design for CB-ACS

Precise channel estimation can be obtained by CB-ACS only when the beam pattern is suitable. On the one hand, the beam pattern should be directional to make the feedback index useful for selecting beam patterns. On the order hand, the pilot beams should be shaped to satisfy the correlation constraints. In this section, based on the two points above, two criteria for the beam pattern are proposed and a suitable beam pattern design is given, followed by the codebook generation.

A. Training Beam Pattern Design

According to the directivity and correlation requirement, taking BS as example, we give two criteria for the pilot beam pattern design.

Criterion 1: For a precoder $\boldsymbol{f}_m\,(m=1,2,\cdots,M)$, the gain function is $C_m(\theta) = \left|\boldsymbol{a}_{\mathrm{T}}^{\mathrm{H}}(\theta)\,\boldsymbol{f}_m\right|$. Its beam coverage, which can be expressed as [7] $\mathcal{CV}(\boldsymbol{f}_m) = \left\{\theta \mid C_m(\theta) \geq \delta \max\limits_{\omega\in[-1,1]}\{C_m(\omega)\}\right\}$, satisfies $\mathcal{CV}(\boldsymbol{f}_m) = \left\{\theta \mid \mathrm{dis}\left(\theta,\bar{\theta}_m,2\right) < 1/M\right\}$, where δ is a parameter within $(0,1)$, $\bar{\theta}_m$ is one of the central angles. In this paper, the set of the central angles is $\left\{\bar{\theta}_m = -1+\frac{2}{M}(m-1), m=1,2,\cdots,M\right\}$.

This criterion comes from the directivity requirement. Under this, M beams uniformly divide the angle domain $[-1,1]$.

Before giving Criterion 2, the spatial correlation is analyzed. If $\mu_{\mathrm{T}}(\theta_{\mathrm{T},1}\theta_{\mathrm{T},2}) = \boldsymbol{a}_{\mathrm{T}}^{\mathrm{H}}(\theta_{\mathrm{T},1})\,\boldsymbol{F}\boldsymbol{F}^{\mathrm{H}}\boldsymbol{a}_{\mathrm{T}}(\theta_{\mathrm{T},2})$ and $\mu_{\mathrm{R}}(\theta_{\mathrm{R},1}\theta_{\mathrm{R},2}) = \boldsymbol{a}_{\mathrm{R}}^{\mathrm{H}}(\theta_{\mathrm{R},1})\,\boldsymbol{W}\boldsymbol{W}^{\mathrm{H}}\boldsymbol{a}_{\mathrm{R}}(\theta_{\mathrm{R},2})$, the spatial correlation function is

$$\begin{aligned}
&\mu\left((\theta_{\mathrm{T},1}\theta_{\mathrm{R},1}),(\theta_{\mathrm{T},2}\theta_{\mathrm{R},2})\right) = \boldsymbol{q}(\theta_{\mathrm{T},1}\theta_{\mathrm{R},1})^{\mathrm{H}}\,\boldsymbol{q}(\theta_{\mathrm{T},2}\theta_{\mathrm{R},2})\\
&=\left[\boldsymbol{F}^{\mathrm{T}}\boldsymbol{a}_{\mathrm{T}}^{*}(\theta_{\mathrm{T},1})\otimes\boldsymbol{W}^{\mathrm{H}}\boldsymbol{a}_{\mathrm{R}}(\theta_{\mathrm{R},1})\right]^{\mathrm{H}}\left[\boldsymbol{F}^{\mathrm{T}}\boldsymbol{a}_{\mathrm{T}}^{*}(\theta_{\mathrm{T},2})\otimes\boldsymbol{W}^{\mathrm{H}}\boldsymbol{a}_{\mathrm{R}}(\theta_{\mathrm{R},2})\right]\\
&\overset{(a)}{=}\left[\boldsymbol{a}_{\mathrm{T}}^{\mathrm{T}}(\theta_{\mathrm{T},1})\,\boldsymbol{F}^{*}\boldsymbol{F}^{\mathrm{T}}\boldsymbol{a}_{\mathrm{T}}^{*}(\theta_{\mathrm{T},2})\right]\left[\boldsymbol{a}_{\mathrm{R}}^{\mathrm{H}}(\theta_{\mathrm{R},1})\,\boldsymbol{W}\boldsymbol{W}^{\mathrm{H}}\boldsymbol{a}_{\mathrm{R}}(\theta_{\mathrm{R},2})\right]\\
&=\mu_{\mathrm{T}}(\theta_{\mathrm{T},1}\theta_{\mathrm{T},2})\,\mu_{\mathrm{R}}(\theta_{\mathrm{R},1}\theta_{\mathrm{R},2}),
\end{aligned} \tag{13}$$

where (a) comes from the formulas $(\boldsymbol{A}\otimes\boldsymbol{B})^{\mathrm{H}} = \boldsymbol{A}^{\mathrm{H}}\otimes\boldsymbol{B}^{\mathrm{H}}$ and $(\boldsymbol{A}\otimes\boldsymbol{B})(\boldsymbol{C}\otimes\boldsymbol{D}) = \boldsymbol{A}\boldsymbol{C}\otimes\boldsymbol{B}\boldsymbol{D}$. Accordingly, we can get the expected spatial correlation function by respectively devising the transmitting and the receiving correlation function. Taking transmitting as example, the second criterion is

Criterion 2: $\forall\theta_{\mathrm{T}} \in [-1,1], \arg\max\limits_{\bar{\theta}_g}\left|\mu_{\mathrm{T}}(\theta_{\mathrm{T}},\bar{\theta}_g)\right| = \arg\min\limits_{\bar{\theta}_g}\mathrm{dis}\left(\theta_{\mathrm{T}},\bar{\theta}_g,2\right)$,where $\bar{\theta}_g$ is one of the predefined grids.

This criterion guarantees that the grid nearest to the AoD is the most likely one to be selected.

The cosine beam pattern, whose ideal gain function is $C_m(\theta) = C\cos\left(\frac{M\pi}{4}\mathrm{dis}\left(\theta,\bar{\theta}_m,2\right)\right)$ when $\mathrm{dis}\left(\theta,\bar{\theta}_m,2\right) \leq \frac{2}{M}$ and $C_m(\theta) = 0$ when $\mathrm{dis}\left(\theta,\bar{\theta}_m,2\right) > \frac{2}{M}\,(m=1,2,\cdots,M)$, satisfies the two criteria. This conclusion will be proved in the next paragraph. An example of the cosine beam pattern is illustrated by the dotted lines in Fig. 4.

For the cosine beam pattern, Criterion 1 is satisfied apparently. Then, we prove that Criterion 2 is also satisfied. If the AOD $\theta_{\mathrm{T}} \in \left[0,\frac{2}{M}\right]$, the correlation function of the ideal beam pattern is $\mu_{\mathrm{T}}(\theta,\theta_{\mathrm{T}}) = C^2\cos\left(\frac{M\pi}{4}(\theta-\theta_{\mathrm{T}})\right)$ when $\theta \in \left[0,\frac{2}{M}\right]$, $\mu_{\mathrm{T}}(\theta,\theta_{\mathrm{T}}) = C^2\cos\left(\frac{M\pi}{4}\theta_{\mathrm{T}}\right)\cos\left(\frac{M\pi}{4}\theta\right)$ when $\theta \in \left[-\frac{2}{M},0\right)$, $\mu_{\mathrm{T}}(\theta,\theta_{\mathrm{T}}) = C^2\sin\left(\frac{M\pi}{4}\theta_{\mathrm{T}}\right)\sin\left(\frac{M\pi}{4}\theta\right)$ when $\theta \in \left(\frac{2}{M},\frac{4}{M}\right]$ and $\mu_{\mathrm{T}}(\theta,\theta_{\mathrm{T}}) = 0$ in other cases. Therefore, $\mu_{\mathrm{T}}(\theta,\theta_{\mathrm{T}})$ decreases with the increase of $\mathrm{dis}(\theta,\theta_{\mathrm{T}},2)$. Since the beams in one level have the same shape and are uniformly distributed in the angle domain, this nature is still available if $\theta_{\mathrm{T}} \in [-1,1]$. Thus, Criteria 2 is satisfied.

Moreover, the cosine beam pattern generates smaller correlation between close AoDs than the rectangular beam pattern in [6]. This brings more accurate estimation of the contributions of the estimated paths in Algorithm 1. It is interesting to further reduce the correlation between close paths.

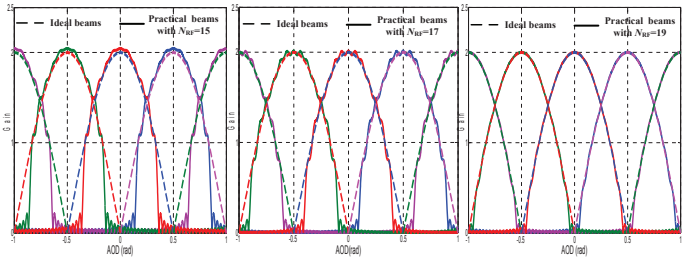

Fig. 4. Beam patterns with $M=4, N_{\mathrm{T}}=64, B=6\mathrm{bit}$.

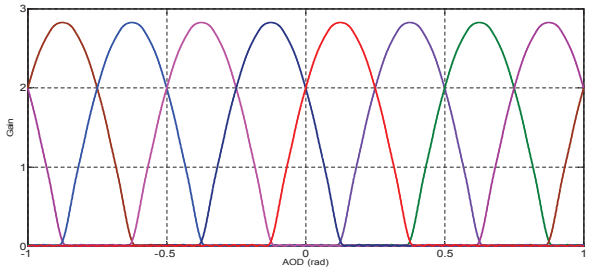

Fig. 5. Beam patterns with $M=8, N_{\mathrm{T}}=64, N_{\mathrm{RF}}=15, B=6\mathrm{bit}$.

B. Hierarchical Codebook Design

According to the ideal gain function obtained in part A, the hierarchical codebook design is described in this part taking

BS as example. Assuming that N_{C} ($N_{\mathrm{C}} \leq N_{\mathrm{RF}}$) RF chains are employed, the RF precoder is $\boldsymbol{F}_{\mathrm{RF}} = [\boldsymbol{w}_1, \boldsymbol{w}_2, \cdots, \boldsymbol{w}_{N_{\mathrm{C}}}]$ and the baseband precoding vector is $\boldsymbol{f}_{\mathrm{BB}} = [b_1, b_2, \cdots, b_{N_{\mathrm{C}}}]^{\mathrm{T}}$. If the gain of $\boldsymbol{w}_n$ is $C_{\boldsymbol{w}_n}(\theta) = \boldsymbol{a}_{\mathrm{T}}^{\mathrm{H}}(\theta)\boldsymbol{w}_n$, the gain function of hybrid precoder is

$$\begin{aligned} C(\theta) &= \left|\boldsymbol{a}_{\mathrm{T}}^{\mathrm{H}}(\theta)\boldsymbol{F}_{\mathrm{RF}}\boldsymbol{f}_{\mathrm{BB}}\right| = \left|\boldsymbol{a}_{\mathrm{T}}^{\mathrm{H}}(\theta)[\boldsymbol{w}_1, \boldsymbol{w}_2, \cdots, \boldsymbol{w}_{N_{\mathrm{C}}}]\boldsymbol{f}_{\mathrm{BB}}\right| \\ &= \left|b_1 C_{\boldsymbol{w}_1}(\theta) + b_2 C_{\boldsymbol{w}_2}(\theta) + \cdots + b_{N_{\mathrm{C}}} C_{\boldsymbol{w}_{N_{\mathrm{C}}}}(\theta)\right|. \quad (14) \end{aligned}$$

The ARVs are taken as the RF precoding vectors. Given the quantization bit B, $\bar{\Phi} = \left\{\bar{\eta}_u \mid \bar{\eta}_u = -1 + \frac{1}{N_B}(u-1), u = 1, 2, \cdots, 2N_B, N_B = 2^B\right\}$ is the collection of quantized angles for phase shifters. One of the RF precoding vectors is

$$\boldsymbol{w}_n = \boldsymbol{a}_{\mathrm{T}}(\xi_n), \quad (15)$$

where $\xi_n \in \bar{\Phi}$ is the central angle of the main lobe generated by $\boldsymbol{w}_n$. If $D_N(\theta) = \frac{\sin\frac{\pi N\theta}{2}}{\sin\frac{\pi\theta}{2}}$, the gain function of $\boldsymbol{w}_n$ is

$$C_{\boldsymbol{w}_n}(\theta) = \left|\boldsymbol{a}_{\mathrm{T}}^{\mathrm{H}}(\theta)\boldsymbol{a}_{\mathrm{T}}(\xi_n)\right| = e^{j\frac{\pi}{2}(N_{\mathrm{T}}-1)(\theta-\xi_n)} D_{N_{\mathrm{T}}}(\theta-\xi_n). \quad (16)$$

In order to design the baseband precoder, we divide $\boldsymbol{f}_{\mathrm{BB}}$ into two parts, namely $\boldsymbol{f}_{\mathrm{BB}} = \boldsymbol{f}_{\mathrm{BB},1} \circ \boldsymbol{f}_{\mathrm{BB},2}$, where

$$\boldsymbol{f}_{\mathrm{BB},1} = \left[e^{j\frac{\pi}{2}(N_{\mathrm{T}}-1)\xi_1}, e^{j\frac{\pi}{2}(N_{\mathrm{T}}-1)\xi_2}, \ldots, e^{j\frac{\pi}{2}(N_{\mathrm{T}}-1)\xi_{N_{\mathrm{C}}}}\right]^{\mathrm{T}} \quad (17)$$

is used to eliminate the phase difference of different RF gain functions in (16) and $\boldsymbol{f}_{\mathrm{BB},2} = [b_{2,1}, b_{2,2}, \cdots, b_{2,N_{\mathrm{C}}}]$. Now the gain function (14) gets

$$\begin{aligned} C(\theta) = |b_{2,1} D_{N_{\mathrm{T}}}(\theta-\xi_1) + b_{2,2} D_{N_{\mathrm{T}}}(\theta-\xi_2) + \\ \cdots + b_{2,N_{\mathrm{C}}} D_{N_{\mathrm{T}}}(\theta-\xi_{N_{\mathrm{C}}})|. \quad (18) \end{aligned}$$

The column vector $\boldsymbol{c}$ is obtained by sampling the expected function $C(\theta)$. The column vector $\boldsymbol{d}(\bar{\eta}_u), \bar{\eta}_u \in \bar{\Phi}$ is obtained by sampling the function $D_{N_{\mathrm{T}}}(\theta - \bar{\eta}_u)$. So $\boldsymbol{c} = \boldsymbol{D}\boldsymbol{f}_{BB,2}$, where $\boldsymbol{D} = [\boldsymbol{d}(\xi_1), \boldsymbol{d}(\xi_2), \cdots, \boldsymbol{d}(\xi_{N_{\mathrm{C}}})]$. Then $\boldsymbol{f}_{BB,2}$ can be obtained by solving

$$\begin{aligned} \boldsymbol{f}_{\mathrm{BB},2}^{\mathrm{opt}} &= \arg\min_{\boldsymbol{D}, \boldsymbol{f}_{\mathrm{BB},2}} \left\|\boldsymbol{c} - \boldsymbol{D}\boldsymbol{f}_{\mathrm{BB},2}\right\|_{\mathrm{F}} \\ \text{s. t. } \boldsymbol{D}_{:,n} &= \boldsymbol{d}(\xi_n), \xi_n \in \bar{\Phi}, n = 1, 2, \cdots, N_{\mathrm{C}}. \quad (19) \end{aligned}$$

The hierarchical codebook generation scheme is given as follow. For each level, two steps are needed.

1)*Basic Code Generation*: In this step, the code whose beam has the central angle 0 is generated. The ideal gain function in this level is $C_s(\theta) = \cos\left(\frac{M_s\pi}{4}\theta\right)$ when $\theta \in \left[-\frac{2}{M_s}, \frac{2}{M_s}\right]$ and $C_s(\theta) = 0$ in other cases, which is sampled as $\boldsymbol{c}_s$. For the symmetry nature of the beam pattern, the number of RF chains $N_{\mathrm{C},s}$ is odd. Inspired by [4], the basic code is designed via OMP as shown in Algorithm 2. In each iteration, if one index is selected, the symmetric one about the central angle is also selected. Note that the index corresponding to the central angle is certain to be selected in the first iteration. With the selected indices, $\boldsymbol{f}_{\mathrm{BB},2}$ is worked out by LS. The output is the basic code $\boldsymbol{f}_{\mathrm{B},s}$.

2)*Beam Rotation*: According to the Theorem 1 in [7], the code $[\boldsymbol{F}_s]_{:,m}$ whose beam has the central angle $\bar{\theta}_m$ is

$$[\boldsymbol{F}_s]_{:,m} = \boldsymbol{f}_{\mathrm{B},s} \circ \frac{1}{\sqrt{N_{\mathrm{T}}}} \boldsymbol{a}_{N_{\mathrm{T}}}(\bar{\theta}_m), \quad (20)$$

where the central angles in the sth level are $\bar{\theta}_m = -1 + 2(m-1)/M_s$, $m = 1, 2, \cdots, M_s$. If M_{P} is even, to make the training beams symmetric about the selected grid, the central angles in the 2nd level to the Sth level are adjusted by adding $1/M_s$, namely $\bar{\theta}_m = -1 + 1/M_s + 2(m-1)/M_s$. Given the accuracy of phase shifters, 2^Bshould be divisible by M_s.

Note that $N_{\mathrm{C},s} = \frac{2N_{\mathrm{T}}}{M_s} + 1$ RF chains is enough for the sth level to generate cosine beams. Thus it is feasible to use partial RF chains. This helps reduce the energy consumption. The examples of beams are shown in Fig. 4 and Fig. 5.

Algorithm 2 Basic code design via OMP

Input: $\boldsymbol{c}_s$, $\bar{\boldsymbol{D}} = [\boldsymbol{d}(\bar{\eta}_1), \boldsymbol{d}(\bar{\eta}_2), \cdots, \boldsymbol{d}(\bar{\eta}_{2N_B})]$, $N_{\mathrm{C},s}$
Initialization: $\boldsymbol{c}_{res} = \boldsymbol{c}_s$, $\mathcal{I}_{\mathrm{d}} = \emptyset$
for $k = 1 : (N_{\mathrm{C},s}+1)/2$ **do**
  $i_1 = \arg\max_u \left(\left[\bar{\boldsymbol{D}}^{\mathrm{T}}\boldsymbol{c}_{res}\right]_u\right)$
  $\mathcal{I}_{\mathrm{d}} = \mathcal{I}_{\mathrm{d}} \cup \{i_1\}$
  $i_2 = \mathrm{cmod}(2(N_B+1) - i_1, 2N_B)$
  $\mathcal{I}_{\mathrm{d}} = \mathcal{I}_{\mathrm{d}} \cup \{i_2\}$
  $\boldsymbol{f}_{\mathrm{BB},2} = \left(\boldsymbol{D}^{\mathrm{H}}\boldsymbol{D}\right)^{-1}\boldsymbol{D}^{\mathrm{H}}\boldsymbol{c}_s$, $\boldsymbol{D} = \left[\bar{\boldsymbol{D}}\right]_{:,\mathcal{I}_{\mathrm{d}}}$
  $\boldsymbol{c}_{\mathrm{res}} = \boldsymbol{c}_s - \boldsymbol{D}\boldsymbol{f}_{\mathrm{BB},2}$
end for
$\mathcal{I}_{\mathrm{d}} = [\mathcal{I}_{\mathrm{d}}]_{2:N_{\mathrm{C},s}}$, $\boldsymbol{D} = \left[\bar{\boldsymbol{D}}\right]_{:,\mathcal{I}_{\mathrm{d}}}$, $\boldsymbol{f}_{\mathrm{BB},2} = \left(\boldsymbol{D}^{\mathrm{H}}\boldsymbol{D}\right)^{-1}\boldsymbol{D}^{\mathrm{H}}\boldsymbol{c}_s$
With the collection of central angles $\bar{\Phi}_{\mathcal{I}_{\mathrm{d}}}$, $\boldsymbol{F}_{\mathrm{RF}}$ and $\boldsymbol{f}_{\mathrm{BB},1}$ are calculated according to equation (15)(17)
Output: $\boldsymbol{f}_{\mathrm{B},s} = \boldsymbol{F}_{\mathrm{RF}}\left(\boldsymbol{f}_{\mathrm{BB},1} \circ \boldsymbol{f}_{\mathrm{BB},2}\right)$, $\boldsymbol{f}_{\mathrm{B},s} = \boldsymbol{f}_{\mathrm{B},s} / \left\|\boldsymbol{f}_{\mathrm{B},s}\right\|_{\mathrm{F}}$

V. Performance Evaluation

Simulation results are given in this section. Consider the system described in Section II. Assume that $N_{\mathrm{T}} = N_{\mathrm{R}} = 64$ antennas and $N_{\mathrm{T}}^{\mathrm{RF}} = N_{\mathrm{R}}^{\mathrm{RF}} = 16$ RF chains are equipped both in BS and MS. The analog phase shifters with the quantization bit $B = 6$ is employed. For the mmWave channel, the AoDs and AoAs are uniformly distributed in $[-1, 1]$. The variance of the complex gain is $\sigma_\alpha^2 = 1$. The codebook in PB-ACS is the hierarchical codebook in [6] with all 16 RF chains employed. For the proposed codebook, the number of RF chains used in the sth levels is $N_{\mathrm{C},s} = \frac{2N_{\mathrm{T}}}{M_s} + 1$ if $N_{\mathrm{C},s} < 16$. Else, $N_{\mathrm{C},s} = 15$. To be fair and get acceptable results, M_1 is set as four in the two codebooks. In addition, M_S is set as N_{T} and G is $4N_{\mathrm{T}}$ in the proposed codebook.

In Fig. 6, the estimation performances with different schemes under the single-path channel model are compared. The performance is evaluated by the average probability of wrong estimation (APWE) defined as $\mathop{\mathbb{E}}_{\theta_{\mathrm{T}}\in[-1,1]}\left[\Pr\left(\mathrm{dis}\left(\hat{\theta}_{\mathrm{T}}, \theta_{\mathrm{T}}, 2\right) > \frac{1}{N_{\mathrm{T}}}\right)\right]$. According to the results, some conclusions can be obtained. First of all, for all the schemes, PA can reduce the APWE effectively. Then, for CB-ACS, the scheme with $K = 4, M_{\mathrm{P}} = 4$ ($T = 48$) and the

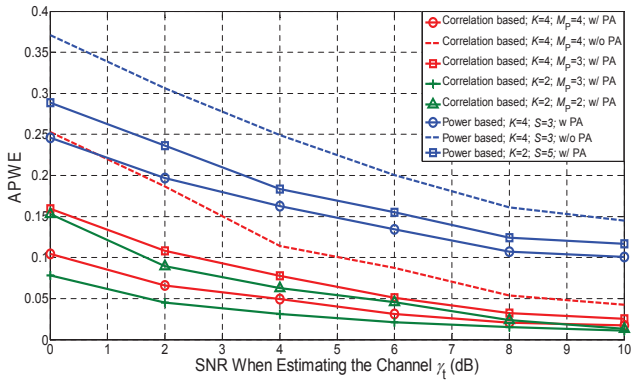

Fig. 6. APWE of different schemes in single-path channel.

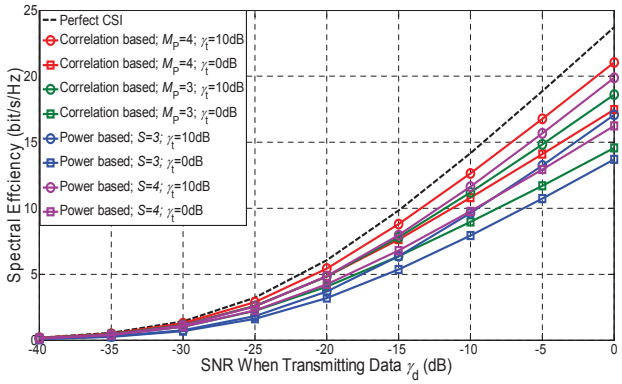

Fig. 7. Achievable Spectral efficiency of different schemes in the channel with $L = 3$ paths.

scheme with $K = 2, M_{\mathrm{P}} = 3\,(T = 52)$ have smaller APWE than the scheme with $K = 4, M_{\mathrm{P}} = 3\,(T = 34)$ and the scheme with $K = 2, M_{\mathrm{P}} = 2\,(T = 32)$. Therefore, the setting with higher training overhead has smaller APWE. In addition, The two PB-ACS schemes with $K = 4, S = 3\,(T = 48)$ and $K = 2, S = 5\,(T = 32)$ have the same training overhead with the CB-ACS schemes with $K = 4, M_{\mathrm{P}} = 4$ and $K = 2, M_{\mathrm{P}} = 2$, respectively. But the CB-ACS schemes have better APWE performance.

After channel estimation, the optimal precoder for communication can be obtained through singular value decomposition (SVD) of the channel matrix [4]. And the optimal combiner is designed by minimizing mean square error (MMSE) [4]. With the optimal precoder/combiner employed, ASE is simulated under the channel with $L = 3$ paths. Assume that the number of paths to be estimated is $L_{\mathrm{d}} = 3$ and the number of data streams is $N_{\mathrm{d}} = L_{\mathrm{d}}$. For each scheme, PA is employed and K is set as 4. The SNR when estimating the channel is $\gamma_{\mathrm{t}} = 10\mathrm{dB}$ and $\gamma_{\mathrm{t}} = 0\mathrm{dB}$. As shown in Fig. 7, The CB-ACS with $M_{\mathrm{P}} = 4$ and the PB-ACS with $S = 3$ have the same training overhead $(T = 144)$, but the former has higher ASE. The ASE of the CB-ACS with $M_{\mathrm{P}} = 3$ is also higher than the PB-ACS with $S = 3$, and it has lower training overhead $(T = 102)$. For the PB-ACS scheme with $S = 4$, the training overhead $(T = 192)$ is raised by adding a level and the estimated angular resolution in theory is same with the CB-ACS schemes, namely $G = 256$. But the CB-ACS scheme with $M_{\mathrm{P}} = 4$ still has higher ASE. As a result, CB-ACS schemes have better ASE performance than PB-ACS even with lower training overhead.

VI. Conclusion

In this paper, the correlation based adaptive compressed sensing method is proposed to estimate the mmWave channel, followed by the codebook design. Simulation results show that under the hardware constraints, the proposed scheme can realize a more precise estimation even with lower training overhead. Thus, larger achievable spectral efficiency can be obtained. For future work, it would be interesting to design the training codebook with less RF chains.

Acknowledgement

The research is sponsored by Project 61471066 supported by NSFC.

References

[1] J. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. Soong and J. Zhang,"What Will 5G Be?" *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 6, pp. 1065-1082, Jun. 2014.

[2] E. G. Larsson, O. Edfors, F. Tufvesson, T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 186-195, Feb. 2014.

[3] F. Sohrabi and W. Y, "Hybrid Digital and Analog Beamforming Design for Large-Scale Antenna Arrays," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 3, pp. 501-513, Apr. 2016.

[4] O.E. Ayach, S. Rajagopal, S. Abusurra, Z. Pi and R.W. Heath, "Spatially Sparse Precoding in Millimeter Wave MIMO Systems," *IEEE Transactions on Wireless Communications*, vol. 13, no. 3, pp. 1499-1513, Mar. 2014.

[5] R. W. Heath, N. G. Prelcic, S. Rangan, W. Roh, and A. Sayeed, "An Overview of Signal Processing Techniques for Millimeter Wave MIMO Systems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 3, pp. 436C453, Apr. 2016.

[6] A. Alkhateeb, O.E. Ayach, G. Leus and R.W. Heath, "Channel Estimation and Hybrid Precoding for Millimeter Wave Cellular Systems," *IEEE Journal of Selected Topics in Signal Processing*, 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 Transactions on Wireless Communications*, vol. 15, no. 5, pp. 3380-3392, May. 2016.

[8] S. Noh, M.D. Zoltowski, and D.J. Love, "Multi-Resolution Codebook Based Beamforming Sequence Design in Millimeter-Wave Systems," *IEEE Global Communications Conference (GLOBECOM)*, Dec. 2015, pp. 1-6.

[9] J. Lee, G. T. Gil and Y. H. Lee, "Channel Estimation via Orthogonal Matching Pursuit for Hybrid MIMO Systems in Millimeter Wave Communications," *IEEE Transactions on Communications*, vol. 64, no. 6, pp. 2370-2386, Jun. 2016.

[10] J. Lee, G. T. Gil and Y. H. Lee, "Exploiting Spatial Sparsity for Estimating Channels of Hybrid MIMO Systems in Millimeter Wave Communications," *IEEE Global Communications Conference (GLOBECOM)*, Dec. 2014, pp. 3326-3331.

[11] T. Cai and L. Wang, "Orthogonal Matching Pursuit for Sparse Signal Recovery With Noise," *IEEE Transactions on Information Theory*, vol. 57, no. 7, pp. 4680-4688, Jul. 2011.

[12] Y. Han and J. Lee, "Two-Stage Compressed Sensing for Millimeter Wave Channel Estimation," *IEEE International Symposium on Information Theory (ISIT)*, Jul. 2016, pp.860-864.