Beam Design with Quantized Phase Shifters for Millimeter Wave Massive MIMO

Kangjian Chen and Chenhao Qi
School of Information Science and Engineering
Southeast University, Nanjing 210096, China
Email: qch@seu.edu.cn

***Abstract*—In this paper, beam design for millimeter wave (mmWave) massive MIMO systems is studied regarding quantized phase shifters and different number of RF chains. Given the objective beam, the beam design problem is formulated as a hybrid optimization problem involving continuous variables as well as discrete variables. To reduce the difficulty in solving this problem, it is converted into several discrete optimization subproblems. Then beam design methods are proposed for the system with only one RF chain and two RF chains, respectively. For the general system with more than two RF chains, the parameter estimation is incorporated into the random search. Based on these findings, a partial random search (PRS) based beam design algorithm is proposed. To further improve the convergence speed, a fast search (FS) based beam design algorithm is proposed. Simulation results verify the effectiveness of the proposed algorithms and show that the beam pattern using the proposed PRS and FS based algorithm can well approach the objective beam with much less RF chains than OMP.**

***Index Terms*—Millimeter wave communication, massive MIMO, beam design, quantized phase shifters**

I. Introduction

Millimeter wave (mmWave) communication shows great promise for future wireless communications due to its rich spectrum resources and high transmission rate [1]–[3]. However, higher carrier frequency leads to larger path loss [4]. To compensate for the path loss, directional transmission based on a large MIMO antenna array, i.e., massive MIMO, is usually adopted. On the other hand, higher carrier frequency makes more antennas integrated into the same area [5].

To make the directional transmission, analog precoding using a phase shifter network is normally employed, where the phase shifters have constant envelop and limited resolution [6]. Since the energy consumption of radio frequency (RF) chains is large, a much smaller number of RF chains than the number of antennas is used in mmWave communications. To achieve parallel data transmission and mitigate the mutual interference among different data steams, digital precoding that is similar as the convention MIMO is employed. In such a hybrid precoding structure including analog precoding and digital precoding, beam training based on codebook is beneficial to acquire the channel state information [7]. To reduce the complexity of beam training, a hierarchical codebook is used in popular, where the wide beams formed by the upper layer codewords covers the narrow beams formed by the lower layer codewords [8]. Several hierarchical codebook design schemes have been proposed [9]–[12]. The beam is designed based on the orthogonal matching pursuit (OMP) algorithm and a phase-shifted discrete Fourier transform (PS-DFT) scheme in [9] and [10], respectively. However, both [9] and [10] need a large number of RF chains when designing high-quality wide beam. In [11], the antennas are divided into several subarrays, where the weighted summation of beams formed by different subarrays is conceived to design the required beam. In [12], the beam design is formulated as an optimization problem, where the ripple in the main and side lobes of the beam is constrained. However, neither [11] nor [12] considers the limited resolution of phase shifters.

In this paper, we consider the beam design with quantized phase shifters and different number of RF chains for mmWave massive MIMO systems. Given the objective beam, we formulate the beam design problem as a hybrid optimization problem involving continuous variables as well as discrete variables. To reduce the difficulty in solving this problem, we convert it into several discrete optimization subproblems. Then we propose beam design methods for the system with only one RF chain and two RF chains, respectively. For the general system with more than two RF chains, we incorporate the parameter estimation into the random search. Based on these findings, we propose a partial random search (PRS) based beam design algorithm. To further improve the convergence speed, we propose a fast search (FS) based beam design algorithm.

The notations are defined as follows. Symbols for matrices (upper case) and vectors(lower case) are in boldface. The set is represented by bold Greek letters. $(\cdot)^*$, $(\cdot)^T$ and $(\cdot)^H$ denote the conjugate, transpose and conjugate transpose (Hermitian) respectively. $[\boldsymbol{A}]_{n,:}$ and $[\boldsymbol{A}]_{:,m}$ denote the nth row and mth column of a matrix $\boldsymbol{A}$ respectively. $[\boldsymbol{a}]_n$, $[\boldsymbol{\Phi}]_n$, $[\boldsymbol{A}]_{n,m}$ denote the nth entry of the vector $\boldsymbol{a}$, the nth entry of the set $\boldsymbol{\Phi}$ and the entry on the nth row mth column of a matrix $\boldsymbol{A}$, respectively. j denotes the square root of -1. In addition, $|\cdot|$ and $\|\cdot\|_2$ denote the absolute value of a scalar and ℓ_2-norm of a vector respectively. $\mathbb{C}$, $\mathbb{R}$ and $\mathbb{E}\{\cdot\}$ denote the set of complex number, the set of real positive number and expectation operation respectively. The complex Gaussian distribution is denoted by $\mathcal{CN}$.

II. Problem Formulation

As shown in Fig. 1, we consider an mmWave massive MIMO system including a transmitter and a receiver equipped

978-1-5386-4727-1/18/$31.00 ©2018 IEEE

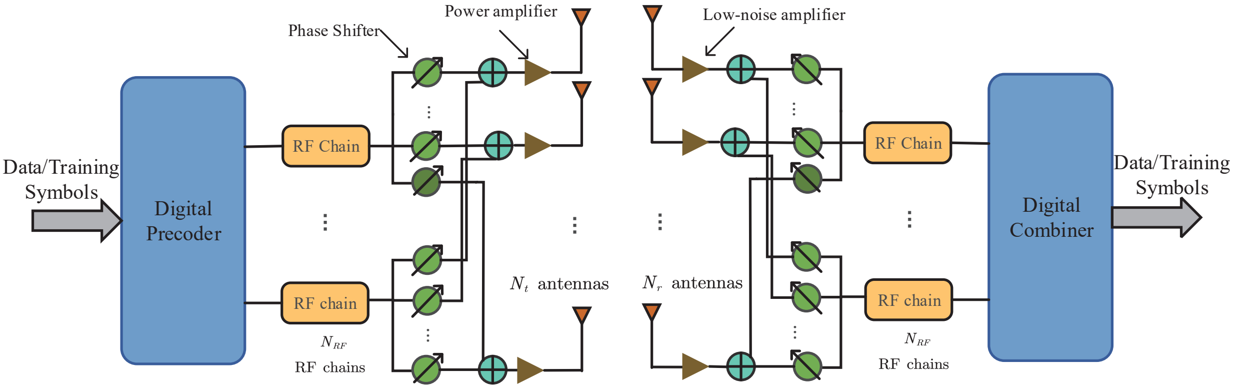

Fig. 1. Illustration of a hybrid precoding and combing structure.

with N_t and N_r antennas, respectively. The antennas at both sides are placed in uniform linear arrays with half wavelength interval. We use the same N_{RF} RF chains at the transmitter and the receiver. At the transmitter, each RF chain is fully connected to N_t antennas via quantized phase shifters, signal combiners and power amplifiers. At the receiver, each RF chain is fully connected to N_r antennas via quantized phase shifters, signal combiners and low-noise amplifiers. Note that the phase shifters usually have limited resolution, e.g., six bits. Without loss of generality, we consider single data stream transmission. Then the received signal after hybrid combining can be expressed as

$$y=\sqrt{P}\boldsymbol{w}_{\mathrm{BB}}^{H}\boldsymbol{W}_{\mathrm{RF}}^{H}\boldsymbol{H}\boldsymbol{F}_{\mathrm{RF}}\boldsymbol{f}_{\mathrm{BB}}x+\boldsymbol{w}_{\mathrm{BB}}^{H}\boldsymbol{W}_{\mathrm{RF}}^{H}\boldsymbol{n} \tag{1}$$

where $\boldsymbol{f}_{\mathrm{BB}}\in\mathbb{C}^{N_{\mathrm{RF}}}$, $\boldsymbol{F}_{\mathrm{RF}}\in\mathbb{C}^{N_t\times N_{\mathrm{RF}}}$, $\boldsymbol{W}_{\mathrm{RF}}\in\mathbb{C}^{N_r\times N_{\mathrm{RF}}}$, $\boldsymbol{w}_{\mathrm{BB}}\in\mathbb{C}^{N_{\mathrm{RF}}}$, $\boldsymbol{n}\in\mathbb{C}^{N_r}$ denote digital precoder, analog precoder, digital combiner, analog combiner, additive white Gaussian noise vector with $\boldsymbol{n}\sim\mathcal{CN}\left(\boldsymbol{0},\sigma_n^2\boldsymbol{I}_{N_r}\right)$. Suppose the total power of the transmitter is P, where the power of the transmit signal x is normalized such that $\mathbb{E}\{xx^*\}=1$ and the hybrid precoder does not provide power gain, i.e., $\|\boldsymbol{F}_{\mathrm{RF}}\boldsymbol{f}_{\mathrm{BB}}\|_2=1$. According to the widely used Saleh-Valenzuela channel model [2], [13], the mmWave MIMO channel matrix $\boldsymbol{H}\in\mathbb{C}^{N_r\times N_t}$ can be formulated as

$$\boldsymbol{H}=\sqrt{\frac{N_tN_r}{L}}\sum_{l=1}^{L}\mu_l\boldsymbol{a}(N_r,\Omega_l^r)\boldsymbol{a}(N_t,\Omega_l^t)^H \tag{2}$$

where L, μ_l, Ω_l^r and Ω_l^t denote the number of multipath, the channel gain, the channel angle-of-arrival (AoA) and channel angle-of-departure (AoD) of the lth path, respectively. In fact, we have $\Omega_l^t=\cos\left(\omega_l^t\right)$ and $\Omega_l^r=\cos\left(\omega_l^r\right)$, where ω_l^t and ω_l^r denote the physical AoA and AoD of the lth path, respectively. Since $\omega_l^t\in[0,2\pi)$ and $\omega_l^r\in[0,2\pi)$, we have $\Omega_l^t\in[-1,1]$ and $\Omega_l^r\in[-1,1]$. The channel steering vector $\boldsymbol{a}$ is defined as

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

where N is the number of antennas, and Ω is the AoA or AoD.

In order to estimate the mmWave MIMO channel, codebook-based beam training is widely adopted [11], [12]. Since the codebook design at the transmitter is similar as that at the receiver, we mainly focus on the codebook design at the transmitter in this paper. Note that the codebook is made up of a number of codewords. The codebook design is essentially the codeword design. Two objectives are commonly used for the codeword design [9], [10], [14]. **1)** If the steering vector $\boldsymbol{a}(N_t,\Omega)$ is covered by the codeword, the absolute beam gain along the direction of Ω is a constant. **2)** If $\boldsymbol{a}(N_t,\Omega)$ is not covered by the codeword, the beam gain is zero. Given Ω, the beam gain of a codeword $\boldsymbol{v}\in\mathbb{C}^{N_t}$ along Ω is defined as

$$A\left(\boldsymbol{v},\Omega\right)=\sum_{n=1}^{N_t}\left[\boldsymbol{v}\right]_n e^{-j\pi(n-1)\Omega}. \tag{4}$$

We denote the beam coverage of the codeword $\boldsymbol{v}$ as $\mathcal{I}_v$. Then the objectives of codeword design are expressed as

$$\left|A\left(\boldsymbol{v},\Omega\right)\right|=\begin{cases}u_v, & \Omega\in\mathcal{I}_v,\\ 0, & \Omega\notin\mathcal{I}_v\end{cases} \tag{5}$$

where u_v is the constant absolute beam gain within the beam coverage. For simplicity, we consider positive beam gain. Then (5) can be rewritten as

$$A\left(\boldsymbol{v},\Omega\right)=\begin{cases}u_v, & \Omega\in\mathcal{I}_v,\\ 0, & \Omega\notin\mathcal{I}_v.\end{cases} \tag{6}$$

We define $\boldsymbol{M}\triangleq[\boldsymbol{a}(N_t,\Omega_1),\boldsymbol{a}(N_t,\Omega_2),\ldots,\boldsymbol{a}(N_t,\Omega_S)]$ as a matrix made up of $S(S>N_t)$ steering vectors, where

$$\Omega_i=-1+(2i-1)/S,\ i=1,2,\ldots,S. \tag{7}$$

We also define a real positive vector $\boldsymbol{u}\in\mathbb{R}^S$, where the $i(i=1,2,\ldots,S)$th entry of $\boldsymbol{u}$ is

$$[\boldsymbol{u}]_i=\begin{cases}u_v, & \Omega_i\in\mathcal{I}_v,\\ 0, & \Omega_i\notin\mathcal{I}_v.\end{cases} \tag{8}$$

Then we have

$$\boldsymbol{M}^H\boldsymbol{v}=\boldsymbol{u}. \tag{9}$$

Note that the rank of $\boldsymbol{M}$ is N_t. Therefore, the least squares (LS) estimation of $\boldsymbol{v}$ is

$$\hat{\boldsymbol{v}}=\left(\boldsymbol{M}\boldsymbol{M}^H\right)^{-1}\boldsymbol{M}\boldsymbol{u}. \tag{10}$$

Usually we normalize $\hat{\boldsymbol{v}}$ by

$$\boldsymbol{v}_{\mathrm{o}} = \frac{\hat{\boldsymbol{v}}}{\|\hat{\boldsymbol{v}}\|_2}. \tag{11}$$

to guarantee each codeword does not provide power gain. (11) is commonly regarded as an objective for the codeword design in the existing literature [7], [9], [15]. The examples of different beam coverage, e.g., $[-1, 0]$, $[0, 0.5]$, and $[0.5, 0.75]$, achieved by $\boldsymbol{v}_{\mathrm{o}}$ are illustrated in Fig. 2, where $N_t = 64$, $S = 2000$ and $u_v = 1$.

The beam design problem is essentially to approach (11) under the constraints of constant envelop and limited resolution of phase shifters, which can be expressed as

$$\begin{aligned}
&\min_{\boldsymbol{F}_{\mathrm{RF}}, \boldsymbol{f}_{\mathrm{BB}}} \ \|\boldsymbol{v}_{\mathrm{o}} - \boldsymbol{F}_{\mathrm{RF}} \boldsymbol{f}_{\mathrm{BB}}\|_2 && \text{(12a)}\\
&\ \ \text{s.t.} \quad \|\boldsymbol{F}_{\mathrm{RF}} \boldsymbol{f}_{\mathrm{BB}}\|_2 = 1, && \text{(12b)}\\
&\qquad\quad [\boldsymbol{F}_{\mathrm{RF}}]_{n,m} = e^{j\delta}, \ \forall \delta \in \boldsymbol{\Phi}_b, && \text{(12c)}\\
&\qquad\quad n = 1, 2, \ldots, N_t, \ m = 1, 2, \ldots, N_{\mathrm{RF}},
\end{aligned}$$

where the constraint of (12b) indicates that the hybrid precoder does not provide power gain, and the constraint of (12c) indicates each entry of the analog precoder satisfies the hardware restrictions of phase shifters. Given the number of bits of quantized phase shifters, denoted as b, all available phase for the quantized phase shifters form a set $\boldsymbol{\Phi}_b$ as

$$\boldsymbol{\Phi}_b = \left[\pi\left(-1+\frac{1}{2^b}\right), \pi\left(-1+\frac{3}{2^b}\right), \cdots \pi\left(1-\frac{1}{2^b}\right)\right]. \tag{13}$$

For simplicity, in (12c) we assume the amplitude of phase shifters is normalized. In practice, we usually first design $\boldsymbol{F}_{\mathrm{RF}}$, based on which we then design $\boldsymbol{f}_{\mathrm{BB}}$ [9]. Therefore, we may temporarily remove (12b) to design $\boldsymbol{F}_{\mathrm{RF}}$, since we can satisfy (12b) by finally adjusting $\boldsymbol{f}_{\mathrm{BB}}$. Then we have

$$\begin{aligned}
&\min_{\boldsymbol{F}_{\mathrm{RF}}, \boldsymbol{f}_{\mathrm{BB}}} \ \|\boldsymbol{v}_{\mathrm{o}} - \boldsymbol{F}_{\mathrm{RF}} \boldsymbol{f}_{\mathrm{BB}}\|_2 && \text{(14a)}\\
&\ \ \text{s.t.} \quad [\boldsymbol{F}_{\mathrm{RF}}]_{n,m} = e^{j\delta}, \ \forall \delta \in \boldsymbol{\Phi}_b, && \text{(14b)}\\
&\qquad\quad n = 1, 2, \ldots, N_t, \ m = 1, 2, \ldots, N_{\mathrm{RF}}.
\end{aligned}$$

Although the OMP algorithm can be used to solve (14) and obtain a well fit for $\boldsymbol{v}_{\mathrm{o}}$, it requires a large number of RF chains [9], which occupies large hardware resource and reduce the energy efficiency.

III. Beam Design

In this paper, we focus on how to solve (14) with smaller number of RF chains compared to OMP. To ease the notation, we denote the $m(m = 1, 2, \ldots, N_{\mathrm{RF}})$th column of $\boldsymbol{F}_{\mathrm{RF}}$ as $\boldsymbol{f}_m$. Then (14) is rewritten as

$$\begin{aligned}
&\min_{\boldsymbol{f}_1, \boldsymbol{f}_2, \ldots, \boldsymbol{f}_{\mathrm{RF}}, \boldsymbol{f}_{\mathrm{BB}}} \ \left\|\boldsymbol{v}_{\mathrm{o}} - \sum_{m=1}^{N_{\mathrm{RF}}} [\boldsymbol{f}_{\mathrm{BB}}]_m \boldsymbol{f}_m\right\|_2 && \text{(15a)}\\
&\quad \text{s.t.} \qquad [\boldsymbol{f}_m]_n = e^{j\delta}, \ \forall \delta \in \boldsymbol{\Phi}_b, && \text{(15b)}\\
&\qquad\qquad\ \ n = 1, 2, \ldots, N_t, \ m = 1, 2, \ldots, N_{\mathrm{RF}}.
\end{aligned}$$

Since each RF chain can support independent transmission of a data stream based on a beam formed by $\boldsymbol{f}_m$, (15) is essentially to find a weighted summation of different beams to approximate $\boldsymbol{v}_{\mathrm{o}}$ given the number of RF chains. However, it is challenging as it is a hybrid optimization problem involving continuous variables $[\boldsymbol{f}_{\mathrm{BB}}]_m, m = 1, 2, \ldots, N_{\mathrm{RF}}$ as well as discrete variables $[\boldsymbol{f}_m]_n, m = 1, 2, \ldots, N_{\mathrm{RF}}; n = 1, 2, \ldots, N_t$. To reduce the difficulty in solving this problem, we set all the continuous variables to be the same, i.e.,

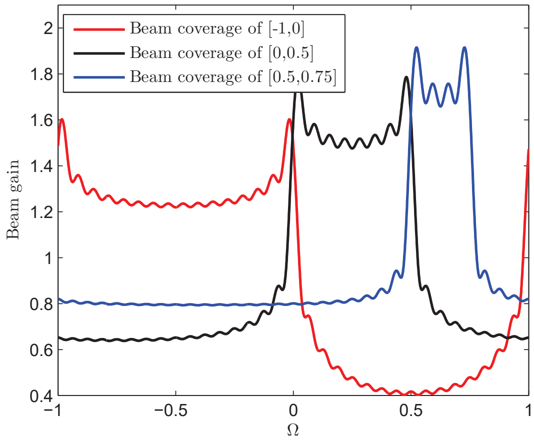

Fig. 2. Different beam coverage for $[-1, 0]$, $[0, 0.5]$ and $[0.5, 0.75]$.

$$\mathbf{f}_{\mathrm{BB}} = \underbrace{[c, c, \cdots, c]}_{N_{\mathrm{RF}}}^T \tag{16}$$

so that we can focus on the discrete optimizations in terms of the quantized phase shifters. Then (15) is converted into a discrete optimization problem as

$$\begin{aligned}
&\min_{\boldsymbol{f}_1, \boldsymbol{f}_2, \ldots, \boldsymbol{f}_{\mathrm{RF}}} \ \left\|\boldsymbol{v}_{\mathrm{o}} - c \sum_{m=1}^{N_{\mathrm{RF}}} \boldsymbol{f}_m\right\|_2 && \text{(17a)}\\
&\quad \text{s.t.} \qquad [\boldsymbol{f}_m]_n = e^{j\delta}, \ \forall \delta \in \boldsymbol{\Phi}_b, && \text{(17b)}\\
&\qquad\qquad\ \ n = 1, 2, \ldots, N_t, \ m = 1, 2, \ldots, N_{\mathrm{RF}}.
\end{aligned}$$

Since $\boldsymbol{v}_{\mathrm{o}}$ is known, we first obtain

$$v_{\max} \triangleq \max_{n=1,2,\cdots,N_t} \left|[\boldsymbol{v}_{\mathrm{o}}]_n\right| \tag{18}$$

showing that the absolute value of each entry of $\boldsymbol{v}_{\mathrm{o}}$ is in the range of $[0, v_{\max}]$. We define $\boldsymbol{f}_{\mathrm{sum}} \triangleq \sum_{m=1}^{N_{\mathrm{RF}}} \boldsymbol{f}_m$. The absolute value of each entry of $c\boldsymbol{f}_{\mathrm{sum}}$ is in the range of $[0, cN_{\mathrm{RF}}]$, where we assume $c > 0$ without loss of generality. We align these two range to minimize the objective shown in (17a) by setting

$$c = \frac{v_{\max}}{N_{\mathrm{RF}}}. \tag{19}$$

Note that the entries of $\boldsymbol{f}_{\mathrm{sum}}$ are mutually independent, implying that the minimization of $\|\boldsymbol{v}_{\mathrm{o}} - c\boldsymbol{f}_{\mathrm{sum}}\|_2$ is essentially the minimization of the absolute value of each entry of $\boldsymbol{v}_{\mathrm{o}} - c\boldsymbol{f}_{\mathrm{sum}}$. Therefore, the optimization problem (17) is divided into N_t subproblems, with the $n(n = 1, 2, \ldots, N_t)$th subproblem expressed as

$$\begin{aligned} \min_{[\boldsymbol{f}_1]_n,[\boldsymbol{f}_2]_n,\cdots,[\boldsymbol{f}_{N_{\mathrm{RF}}}]_n} & \left| \frac{1}{c}\left[\boldsymbol{v}_{\mathrm{o}}\right]_n - \sum_{m=1}^{N_{\mathrm{RF}}} \left[\boldsymbol{f}_m\right]_n \right| \\ \text{s.t.} \qquad & \left[\boldsymbol{f}_m\right]_n = e^{j\delta},\ \forall\delta\in\boldsymbol{\Phi}_b, \\ & m=1,2,\cdots,N_{\mathrm{RF}}. \end{aligned} \tag{20}$$

To ease the notation, we define

$$\alpha_n e^{j\beta_n} \triangleq \frac{1}{c}\left[\boldsymbol{v}_{\mathrm{o}}\right]_n, \tag{21}$$

where $\alpha_n \in [0, N_{\mathrm{RF}}]$ and $\beta_n \in [-\pi,\pi)$ denote the amplitude and the phase, respectively. Since the method to solve (20) is exactly the same for different n, we can omit the subscript n and define $e^{j\theta_m} \triangleq [\boldsymbol{f}_m]_n$. Then (20) is rewritten as

$$\min_{\theta_1,\theta_2,\cdots,\theta_{N_{\mathrm{RF}}}} \left| \alpha_n e^{j\beta_n} - \sum_{m=1}^{N_{\mathrm{RF}}} e^{j\theta_m} \right| \tag{22a}$$

$$\text{s.t.} \qquad \theta_m \in \boldsymbol{\Phi}_b,\ m=1,2,\cdots,N_{\mathrm{RF}}. \tag{22b}$$

To solve (22), we consider three different cases, including $N_{\mathrm{RF}}=1$, $N_{\mathrm{RF}}=2$ and $N_{\mathrm{RF}}>2$.

A. $N_{\mathrm{RF}}=1$

In case of $N_{\mathrm{RF}}=1$, there is only one variable θ_1 in (22), which can be rewritten as

$$\min_{\theta_1\in\boldsymbol{\Phi}_b} \left|\alpha_n e^{j\beta_n} - e^{j\theta_1}\right|. \tag{23}$$

The solution to (23) is denoted as $\widetilde{\theta}_1$, which is essentially to find an entry from $\boldsymbol{\Phi}_b$ closest to β_n.

B. $N_{\mathrm{RF}}=2$

In case of $N_{\mathrm{RF}}=2$, there are two variables θ_1 and θ_2 in (22), which can be rewritten as

$$\min_{\theta_1\in\boldsymbol{\Phi}_b,\ \theta_2\in\boldsymbol{\Phi}_b} \left|\alpha_n e^{j\beta_n} - e^{j\theta_1} - e^{j\theta_2}\right|. \tag{24}$$

We first neglect the constrains of $\theta_1 \in \boldsymbol{\Phi}_b$ and $\theta_2 \in \boldsymbol{\Phi}_b$ to solve

$$\min_{\theta_1,\theta_2} \left|\alpha_n e^{j\beta_n} - e^{j\theta_1} - e^{j\theta_2}\right|. \tag{25}$$

Supposing $\alpha_n e^{j\beta_n} = e^{j\theta_1} + e^{j\theta_2}$, where $\alpha_n \in [0,2]$ according to (21), we have

$$\begin{cases} \cos(\theta_1-\beta_n)+\cos(\theta_2-\beta_n) & =\alpha_n, \\ \sin(\theta_1-\beta_n)+\sin(\theta_2-\beta_n) & =0. \end{cases} \tag{26}$$

The solutions to (26) are denoted as $\bar{\theta}_1$ and $\bar{\theta}_2$. Then we have

$$\begin{cases} \bar{\theta}_1 = \beta_n + \arccos(\frac{\alpha_n}{2}), \\ \bar{\theta}_2 = \beta_n - \arccos(\frac{\alpha_n}{2}), \end{cases} \tag{27}$$

or

$$\begin{cases} \bar{\theta}_1 = \beta_n - \arccos(\frac{\alpha_n}{2}), \\ \bar{\theta}_2 = \beta_n + \arccos(\frac{\alpha_n}{2}). \end{cases} \tag{28}$$

The solutions to (24) are

$$\begin{cases} \widetilde{\theta}_1 = \arg\min\limits_{\theta_1\in\boldsymbol{\Phi}_b} |\theta_1-\bar{\theta}_1|, \\ \widetilde{\theta}_2 = \arg\min\limits_{\theta_2\in\boldsymbol{\Phi}_b} |\theta_2-\bar{\theta}_2|, \end{cases} \tag{29}$$

which is essentially to find two entries from $\boldsymbol{\Phi}_b$ closest to $\bar{\theta}_1$ and $\bar{\theta}_2$, respectively.

Algorithm 1 PRS-based Beam Design Algorithm

```
**Input:** $N_{\mathrm{RF}}$, $b$, $d$, $\boldsymbol{v}_{\mathrm{o}}$, $K_{\max}$.
Obtain $\boldsymbol{\Phi}_b$ and $\boldsymbol{\Phi}_d$ via (13), respectively.
Obtain $v_{\max}$ and $c$ via (18) and (19), respectively.
**for** $n=1,2,\ldots,N_t$ **do**
    Obtain $\alpha_n$ and $\beta_n$ via (21).
    **if** $N_{\mathrm{RF}}=1$ **then**
        Obtain $\widetilde{\theta}_1$ via (23).
    **else if** $N_{\mathrm{RF}}=2$ **then**
        Obtain $\widetilde{\theta}_1$ and $\widetilde{\theta}_2$ via (29).
    **else**
        Set $k=1$.
        **while** $k \le K_{\max}$ **do**
            Randomly select $\hat{\theta}_3,\cdots,\hat{\theta}_{\mathrm{RF}}$ from $\boldsymbol{\Phi}_d$.
            Obtain $\hat{\theta}_1$ and $\hat{\theta}_2$ via (34).
            Compute $g(\hat{\theta}_1,\hat{\theta}_2,\cdots,\hat{\theta}_{N_{\mathrm{RF}}})$ via (35).
            $k \leftarrow k+1$.
        **end while**
        Obtain $\widetilde{\theta}_1,\widetilde{\theta}_2,\cdots,\widetilde{\theta}_{N_{\mathrm{RF}}}$.
    **end if**
    Obtain $[\widetilde{\mathbf{F}}_{\mathrm{RF}}]_{:,n}$ via (36).
**end for**
Obtain $\widetilde{\boldsymbol{f}}_{\mathrm{BB}}$ via (37).
**Output:** $\widetilde{\mathbf{F}}_{\mathrm{RF}}$ and $\widetilde{\boldsymbol{f}}_{\mathrm{BB}}$.
```

C. $N_{\mathrm{RF}}>2$

In case of $N_{\mathrm{RF}}>2$, there are more than two variables in (22). If we first neglect the constrains of (22b) to solve (22a), just as the case of $N_{\mathrm{RF}}=2$, the problem will be underdetermined, where the unknown variables are more than the equations. In this context, one method is to use the random search (RS), which repeatedly selects N_{RF} entries from $\boldsymbol{\Phi}_b$ as the value of $\theta_1,\theta_2,\ldots,\theta_{N_{\mathrm{RF}}}$ and figures out the objective of $|\alpha_n e^{j\beta_n} - \sum_{m=1}^{N_{\mathrm{RF}}} e^{j\theta_m}|$, and finally outputs the combination of $\theta_1,\theta_2,\ldots,\theta_{N_{\mathrm{RF}}}$ achieving the minimal objective. However, the computational complexity is very high and the convergence speed is too slow for such method.

Motivated by the case of $N_{\mathrm{RF}}=2$, we incorporate the calculation of (29) into the RS method and propose a partial random search (PRS) method. For each computation of the objective of $|\alpha_n e^{j\beta_n} - \sum_{m=1}^{N_{\mathrm{RF}}} e^{j\theta_m}|$, we randomly select $N_{\mathrm{RF}}-2$ entries from $\boldsymbol{\Phi}_b$ as the value of $\theta_3,\ldots,\theta_{N_{\mathrm{RF}}}$, while we use (29) to calculate θ_1 and θ_2. Since θ_1 and θ_2 are calculated with the purpose to minimize the objective instead of randomly selected as in RS, we may increase the convergence speed for PRS. In fact, we may use a smaller number of bits than that of the phase shifters, i.e., $d(d \le b)$, so that the search space for $\theta_3,\ldots,\theta_{N_{\mathrm{RF}}}$ can be smaller and therefore the computational complexity is lower. To be specific, in the PRS method we iteratively perform the following three steps as

1) We randomly select $N_{\mathrm{RF}}-2$ entries from $\boldsymbol{\Phi}_d$, supposed to be $\hat{\theta}_3,\cdots,\hat{\theta}_{N_{\mathrm{RF}}}$. Note that these $N_{\mathrm{RF}}-2$ entries can also be the same.

2) Similar as the case of $N_{\rm RF} = 2$, we compute θ_1 and θ_2 by

$$\min_{\theta_1,\theta_2} \left| \gamma_n e^{j\phi_n} - e^{j\theta_1} - e^{\theta_2} \right| \tag{30}$$

where

$$\gamma_n e^{j\phi_n} \triangleq \alpha_n e^{j\beta_n} - \sum_{m=3}^{N_{\rm RF}} e^{j\hat{\theta}_m} \tag{31}$$

with $\gamma_n \in [0, 2N_{\rm RF} - 2]$ and $\phi_n \in [-\pi, \pi)$ representing the amplitude and the phase, respectively. The solutions to (30) are denoted as $\bar{\theta}_1$ and $\bar{\theta}_2$. Then we have

$$\begin{cases} \bar{\theta}_1 = \phi_n + \arccos(\frac{\gamma_n}{2}), \\ \bar{\theta}_2 = \phi_n - \arccos(\frac{\gamma_n}{2}), \end{cases} \tag{32}$$

or

$$\begin{cases} \bar{\theta}_1 = \phi_n - \arccos(\frac{\gamma_n}{2}), \\ \bar{\theta}_2 = \phi_n + \arccos(\frac{\gamma_n}{2}), \end{cases} \tag{33}$$

where $\arccos(\cdot)$ denotes the complex-valued arc-cosine function in this case. We find two entries $\hat{\theta}_1$ and $\hat{\theta}_2$ from $\mathbf{\Phi}_b$ closest to $\bar{\theta}_1$ and $\bar{\theta}_2$, respectively, as follows

$$\begin{cases} \hat{\theta}_1 = \arg \min\limits_{\theta_1 \in \mathbf{\Phi}_b} |\theta_1 - \bar{\theta}_1|, \\ \hat{\theta}_2 = \arg \min\limits_{\theta_2 \in \mathbf{\Phi}_b} |\theta_2 - \bar{\theta}_2|. \end{cases} \tag{34}$$

3) With the obtained $\hat{\theta}_3, \cdots, \hat{\theta}_{N_{\rm RF}}$ in step 1) and $\hat{\theta}_1, \hat{\theta}_2$ in step 2), we figure out the objective by

$$g(\hat{\theta}_1, \hat{\theta}_2, \cdots, \hat{\theta}_{N_{\rm RF}}) = \left| \alpha_n e^{j\beta_n} - \sum_{m=1}^{N_{\rm RF}} e^{j\hat{\theta}_m} \right|. \tag{35}$$

We iteratively run the above three steps until the pre-specified maximal number of iteration $K_{\max}$ is reached. The combination of $\hat{\theta}_1, \hat{\theta}_2, \cdots, \hat{\theta}_{N_{\rm RF}}$ achieving the minimal $g(\hat{\theta}_1, \hat{\theta}_2, \cdots, \hat{\theta}_{N_{\rm RF}})$ during the iterations is denoted as $\widetilde{\theta}_1, \widetilde{\theta}_2, \cdots, \widetilde{\theta}_{N_{\rm RF}}$.

The proposed PRS-based beam design algorithm is shown in **Algorithm 1**. Given the $n (n = 1, 2, \ldots, N_t)$th entry of $\boldsymbol{v}_{\rm o}$, we set the nth row the designed $\mathbf{F}_{\rm RF}$ as

$$[\widetilde{\mathbf{F}}_{\rm RF}]_{n,:} = \left[e^{j\widetilde{\theta}_1}, e^{j\widetilde{\theta}_2}, \cdots, e^{j\widetilde{\theta}_{N_{\rm RF}}} \right]. \tag{36}$$

After N_t iterations, we output $\widetilde{\mathbf{F}}_{\rm RF}$. According to (16) and (12b), we have

$$\widetilde{\boldsymbol{f}}_{\rm BB} = \frac{\boldsymbol{f}_{\rm BB}}{\|\widetilde{\boldsymbol{F}}_{\rm RF} \boldsymbol{f}_{\rm BB}\|_2}. \tag{37}$$

In the PRS-based beam design algorithm, the convergence speed is slow due to the manner of random search. To improve the convergence speed, a fast search (FS)-based beam design algorithm summarized in **Algorithm 2** is proposed.

Algorithm 2 FS-based Beam Design Algorithm

```
1: Input: $N_{\rm RF}$, $b$, $d$, $\boldsymbol{v}_{\rm o}$.
2: Obtain $\mathbf{\Phi}_b$ and $\mathbf{\Phi}_d$ via (13), respectively.
3: Obtain $v_{\max}$ and $c$ via (18) and (19), respectively.
4: for $n = 1, 2, \ldots, N_t$ do
5:     Obtain $\alpha_n$ and $\beta_n$ via (21).
6:     if $N_{\rm RF} = 1$ then
7:         Obtain $\widetilde{\theta}_1$ via (23).
8:     else if $N_{\rm RF} = 2$ then
9:         Obtain $\widetilde{\theta}_1$ and $\widetilde{\theta}_2$ via (29).
10:    else
11:        Set $k = 1$.
12:        Initialize $\theta_3^0, \theta_4^0, \cdots, \theta_{N_{\rm RF}}^0$.
13:        while (41) is not satisfied do
14:            Obtain $\theta_3^k, \theta_4^k, \cdots, \theta_{N_{\rm RF}}^k$ via (39) and (40).
15:            $k \leftarrow k + 1$.
16:        end while
17:        Obtain $\widetilde{\theta}_1, \widetilde{\theta}_2, \ldots, \widetilde{\theta}_{N_{\rm RF}}$ via (42).
18:    end if
19:    Obtain $[\widetilde{\mathbf{F}}_{\rm RF}]_{:,n}$ via (36).
20: end for
21: Obtain $\widetilde{\boldsymbol{f}}_{\rm BB}$ via (37).
22: Output: $\widetilde{\mathbf{F}}_{\rm RF}$ and $\widetilde{\boldsymbol{f}}_{\rm BB}$.
```

In the case of $N_{\rm RF} > 2$, we first randomly select $N_{\rm RF} - 2$ entries from $\mathbf{\Phi}_d$ as the values of $\theta_3, \cdots, \theta_{N_{\rm RF}}$, denoted by $\theta_3^0, \cdots, \theta_{N_{\rm RF}}^0$, where the superscript "0" represents the number of iterations. At the $k (k \geq 1)$th iteration, we determine θ_p^k, where

$$p = \operatorname{mod}(k - 1, N_{\rm RF} - 2) + 3 \tag{38}$$

as follows. We keep all the entries except θ_p^k to be the same as those in the $(k-1)$th iteration, which is expressed as

$$\theta_m^k = \theta_m^{k-1}, m = 3, \cdots, N_{\rm RF}, m \neq p. \tag{39}$$

Then given these $N_{\rm RF} - 3$ entries, we test all entries from $\mathbf{\Phi}_d$ to determine θ_p^k. For each test, given an entry from $\mathbf{\Phi}_d$, denoted as $\hat{\theta}_p \in \mathbf{\Phi}_d$, we obtain $\hat{\theta}_1$ and $\hat{\theta}_2$ via (34) and then compute $g(\hat{\theta}_1, \hat{\theta}_2, \theta_3^k, \cdots, \theta_{p-1}^k, \hat{\theta}_p, \theta_{p+1}^k, \ldots, \theta_{N_{\rm RF}}^k)$. From all of these tests, we find a best $\hat{\theta}_p \in \mathbf{\Phi}_d$ satisfying

$$\min_{\hat{\theta}_p \in \mathbf{\Phi}_d} g(\hat{\theta}_1, \hat{\theta}_2, \theta_3^k, \cdots, \theta_{p-1}^k, \hat{\theta}_p, \theta_{p+1}^k, \ldots, \theta_{N_{\rm RF}}^k). \tag{40}$$

The solution to (40) is denoted as θ_p^k.

We iteratively perform these steps until the stop condition expressed as

$$\theta_m^k = \theta_m^{k+3-N_{\rm RF}}, \ m = 3, 4, \ldots, N_{\rm RF} \tag{41}$$

is satisfied. In fact, (41) means the exactly same routine of the iterations is repeated again, which indicates the results thereafter will keep the same.

As an example, in Fig. 3 we illustrate the running process from step 11 to step 16 of **Algorithm 2**, where $b = 5$, $d = 4$ and $N_{\rm RF} = 5$. The numbers in boxes represent the indices of the selected entries from $\mathbf{\Phi}_d$. Firstly, θ_3, θ_4 and θ_5 are randomly initialized, e.g., θ_3^0, θ_4^0 and θ_5^0 are the 7th, 2nd and 3rd entry of $\mathbf{\Phi}_d$, respectively. At the first iteration ($k = 1$), we fix $\theta_4^1 = \theta_4^0$ and $\theta_5^1 = \theta_5^0$ to find θ_3^1. We test all possibilities of θ_3^1. For each test, given an entry from $\mathbf{\Phi}_d$, denoted as $\hat{\theta}_3 \in \mathbf{\Phi}_d$, we obtain $\hat{\theta}_1$ and $\hat{\theta}_2$ via (34) and then compute the objective $g(\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3, \theta_4^1, \theta_5^1)$. Suppose from all the tests

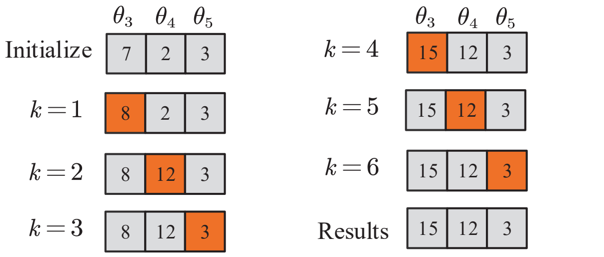

Fig. 3. Illustration of the running process for some steps of Algorithm 2.

we obtain θ_3^1 to be the 8th entry of $\mathbf{\Phi}_d$, which achieves the minimal objective for the first iteration. At the second iteration ($k=2$), we fix $\theta_3^2 = \theta_3^1$ and $\theta_5^2 = \theta_5^1$ and repeat the same process to determine θ_4^2, where θ_4^2 is supposed to be the 12rd entry of $\mathbf{\Phi}_d$. We continue the iterations for $k=3,4,5,6$. Note that $\theta_3^6 = \theta_3^4, \theta_4^6 = \theta_4^4, \theta_5^6 = \theta_5^4$ satisfying (41), which shows the results keep the same for three consecutive iterations, we stop the iterations. It is worth to mention that although the results keep the same for $k=2$ and $k=3$, the results for $k=4$ may be different and therefore the iteration can not be stopped.

Finally we obtain

$$\widetilde{\theta}_m = \theta_m^k, \ m = 1,2,\ldots,N_{\rm RF}. \qquad (42)$$

Given the $n(n=1,2,\ldots,N_t)$th entry of $\boldsymbol{v}_{\rm o}$, $[\mathbf{F}_{\rm RF}]_{n,:}$ can be obtained via (36). Then we can obtain $\widetilde{\boldsymbol{f}}_{\rm BB}$ via (37).

IV. Simulation Results

Now we evaluate the performance of the proposed beam design algorithms.

In Fig. 4, we compare the beam pattern using the OMP, PRS and FS based beam design algorithms. We set $N_t = 32$ and $\mathcal{I}_v = [-1, 0]$. We set $N_{\rm RF} = 4$, $b = 6$ and $d = 5$ for the PRS and FS based beam design algorithms. For PRS, we set $K_{\max} = 100$. For OMP, we set $N_{\rm RF} = 4$ and $N_{\rm RF} = 15$, respectively, while the number of bits of quantized phase shifters is fixed to be six. Given the objective beam generated via (11), we use the OMP, PRS and FS based beam design algorithms to approach the objective beam. It is seen that the beam patterns designed by PRS and FS can well approach the objective beam, while OMP using both $N_{\rm RF} = 4$ and $N_{\rm RF} = 15$ RF chains has large deviation. In fact, we find it requires $N_{\rm RF} = 28$ RF chains for OMP to get a well approximated beam pattern. Therefore, the proposed PRS and FS based beam design algorithms can substantially reduce the number of RF chains, which is significant in saving the hardware resource for practical mmWave massive MIMO systems.

In Fig. 5, we compare the convergence speed for the RS, PRS and FS based beam design algorithms in terms of the number of computation of the objective in (35). We set $N_t = 32$, $\mathcal{I}_v = [-0.5, 0]$, $b = 5$ and $d = 4$. It is seen that the convergence speed of FS is faster than the other algorithms. Under the constrains of the same number of computation of

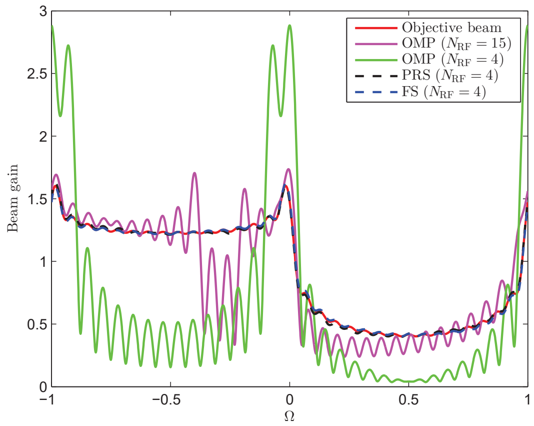

Fig. 4. Comparisons of beam patterns for the OMP, PRS and FS based beam design algorithms.

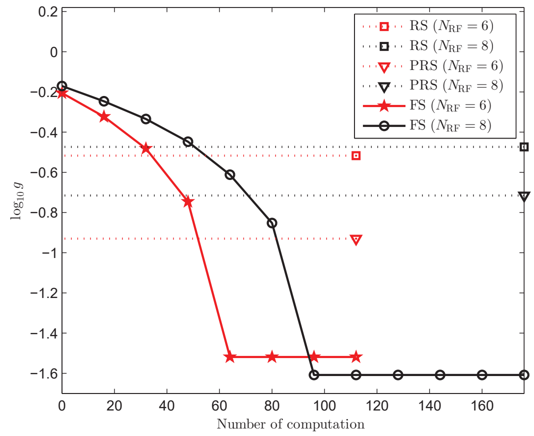

Fig. 5. Comparisons of convergence speed for the RS, PRS and FS based beam design algorithms.

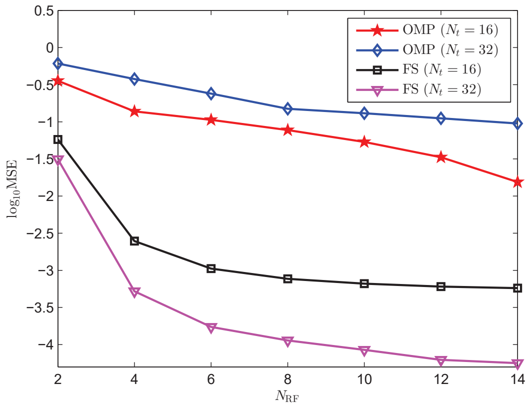

Fig. 6. Comparisons of MSE for the OMP and FS based beam design algorithms.

the objective which essentially corresponds to the hardware constrains, FS can achieve the smaller objective than the other algorithms. In particular, FS with $N_{\rm RF} = 6$ is faster convergent than that with $N_{\rm RF} = 8$, since less variables are involved in $N_{\rm RF} = 6$ than $N_{\rm RF} = 8$. However, given enough running time, FS with $N_{\rm RF} = 8$ achieves the smaller objective than that with $N_{\rm RF} = 6$.

In Fig. 6, we compare the mean squared error (MSE) for the OMP and FS based beam design algorithms. We set $\mathcal{I}_v = [-0.5, 0]$. For $N_t = 16$, we set the number of bits of quantized phase shifters to be four for OMP, while we set $b = 4$ and $d = 3$ for FS. For $N_t = 32$, we set the number of bits of quantized phase shifters to be five for OMP, while we set $b = 5$ and $d = 4$ for FS. It is seen that the MSE of FS is much smaller than that of OMP. The MSE of both FS and OMP gets smaller as $N_{\rm RF}$ increases. In addition, with the increase of N_t and the number of bits of quantized phase shifters, FS performs better owing to the improved resolution of phase shifters, while OMP performs worse since the number of unknown variables to be estimated increases. Therefore, FS is more appropriate for large number of antennas than OMP.

V. Conclusion

We have considered the beam design with quantized phase shifters and different number of RF chains for mmWave massive MIMO systems. Given the objective beam, we have formulated the beam design problem as a hybrid optimization problem and then converted it into several discrete optimization subproblems. We have proposed a PRS based beam design algorithm and a FS based beam design algorithm. Future work will focus on the beam training using the designed beam with quantized phase shifters.

Acknowledgment

This work is supported in part by the National Natural Science Foundation of China (NSFC) under Grants 61871119 and 61302097 and by the Natural Science Foundation of Jiangsu Province under Grant BK20161428.

References

[1] A. Thornburg, T. Bai, and R. W. Heath, "Performance analysis of outdoor mmWave Ad Hoc networks," *IEEE J. Sel. Top. Signal Process.*, vol. 64, no. 15, pp. 4065–4079, Aug. 2016.

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

[3] W. Ma and C. Qi, "Beamspace channel estimation for millimeter wave massive MIMO system with hybrid precoding and combining," *IEEE Trans. Signal Process.*, accepted, 2018.

[4] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, Jan. 2013.

[5] W. Hong, K.-H. Baek, Y. Lee, Y. Kim, and S.-T. Ko, "Study and prototyping of practically large-scale mmWave antenna systems for 5G cellular devices," *IEEE Commun. Mag.*, vol. 52, no. 9, pp. 63–69, Sep. 2014.

[6] R. Mndez-Rial, C. Rusu, N. Gonzlez-Prelcic, A. Alkhateeb, and R. W. Heath, "Hybrid MIMO architectures for millimeter wave communications: Phase shifters or switches?" *IEEE Access*, vol. 4, pp. 247–267, Jan. 2016.

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

[8] 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.

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

[10] S. Noh, M. D. Zoltowski, and D. J. Love, "Multi-resolution codebook based beamforming sequence design in millimeter-wave systems," in *2015 IEEE Global Commun. Conf. (GLOBECOM)*, San Diego, CA, USA, Dec. 2015, pp. 1–6.

[11] 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.

[12] J. Zhang, Y. Huang, Q. Shi, J. Wang, and L. Yang, "Codebook design for beam alignment in millimeter wave communication systems," *IEEE Trans. Commun.*, vol. 65, no. 11, pp. 4980–4995, Nov. 2017.

[13] W. Ma and C. Qi, "Channel estimation for 3-D lens millimeter wave massive MIMO system," *IEEE Commun. Lett.*, vol. 21, no. 9, pp. 2045–2048, Sep. 2017.

[14] Y. Sun and C. Qi, "Weighted sum-rate maximization for analog beamforming and combining in millimeter wave massive MIMO communications," *IEEE Commun. Lett.*, vol. 21, no. 8, pp. 1883–1886, Aug. 2017.

[15] H. Seleem, A. I. Sulyman, and A. Alsanie, "Hybrid precoding-beamforming design with Hadamard RF codebook for mmWave large-scale MIMO systems," *IEEE Access*, vol. 5, pp. 6813–6823, Jun. 2017.