A LOW-COMPLEXITY BEAMFORMING METHOD BY ORTHOGONAL CODEBOOKS FOR MILLIMETER WAVE LINKS

Hsiao-Lan Chiang, Wolfgang Rave, Tobias Kadur, and Gerhard Fettweis

Technische Universität Dresden, Dresden, Germany
Email: {hsiao-lan.chiang, rave, tobias.kadur, gerhard.fettweis}@ifn.et.tu-dresden.de

ABSTRACT

Combining analog with digital beamforming in the sense of hybrid beamforming is one of the promising solutions to maximize throughput for millimeter wave links. Due to a concern of complexity, codebooks used in the analog beamforming are limited to their size. This paper shows that if the limited codebooks are made up of orthogonal steering vectors, the analog beamforming can be implemented with low complexity by exploiting *implicit* knowledge of the channel and then the resulting effective channel should be estimated *explicitly* to determine the optimal weighting coefficients in the digital beamforming. The simulation results show that the proposed low-complexity beamforming method can achieve nearly the same data rates as the one with perfectly known channel state information.

***Index Terms*—** millimeter wave, orthogonal codebook, hybrid beamforming, analog beam selection, orthogonal matching pursuit.

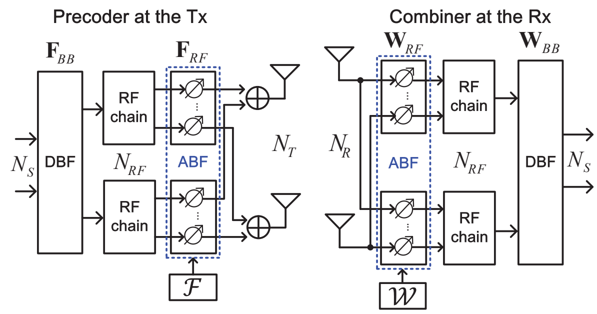

Fig. 1. The beamforming system diagram ($N_S \leq N_{RF} \ll N_T$ (or N_R)). Each RF chain connects to multiple phase shifters and includes a digital-to-analog converter (DAC) at the transmitter or an analog-to-digital converter (ADC) at the receiver.

1. INTRODUCTION

To overcome the severe path loss at millimeter wave (mmWave) systems [1][2], hybrid beamforming is one of the popular approaches to achieve high antenna gains and meanwhile to support spatial multiplexing by a suitable digital combination of the baseband signals in the sense of digital beamforming [3][4][5]. The standard solution to the hybrid beamformer (HBF) reconstruction is to minimize the Frobenius norm of the error between the reconstructed one and the right (or left) singular vectors of the channel matrix [3] and can be obtained by the orthogonal matching pursuit (OMP), which is a greedy algorithm to select a column that most correlated with the current residual in every iteration [6].

One of the challenges of the HBF reconstruction is the computational complexity of the high-dimensional singular value decomposition (SVD) of the channel matrix $\mathbf{H}$. To simplify the procedure, our previous work in [7] presents that the SVD of the estimate of $\mathbf{H}$ can be omitted if the selected array response vectors are orthogonal. In this paper we extend this idea by the following intuitive approach: recognizing that only the *implicit* knowledge of $\mathbf{H}$ is necessary for the analog beamformer (ABF) reconstruction (i.e., neither the precise channel state information (CSI) nor the underlying angles of arrival and departure (AoAs, AoDs) has to be aware of). Such information can be obtained by estimating the coupling between the transmitter and receiver under the given codebook constraints, and then the combination of the analog beams yielding the maximum power of the coupling leads to the optimal solution. Subsequently the resulting effective channel matrix of the order of the number of data streams should be estimated *explicitly* to determine the optimal weighting coefficients in the digital beamformer (DBF).

We use the following notations throughout this paper. a is a scalar, $\mathbf{a}$ is a column vector, and $\mathbf{A}$ is a matrix. a_i denotes the i^{th} entry of $\mathbf{a}$; $a_{i,j}$ denotes the $(i,j)^{\text{th}}$ entry of $\mathbf{A}$; $\mathbf{a}(i)$ denotes the i^{th} column vector of $\mathbf{A}$; $\mathbf{A}(1:N)$ denotes the first N column vectors of $\mathbf{A}$; $\mathbf{A}(1:N,1:N)$ denotes the $N\times N$ submatrix extracted from the upper-left corner of $\mathbf{A}$. $\mathbf{A}^*$, $\mathbf{A}^H$, and $\mathbf{A}^T$ denote the complex conjugate, Hermitian transpose and transpose of $\mathbf{A}$. $\|\mathbf{a}\|_0$ and $\|\mathbf{a}\|_2$ denote the l_0-norm [8] and 2-norm of $\mathbf{a}$; $|\mathbf{A}|$ and $\|\mathbf{A}\|_F$ denote the determinant and Frobenius norm of $\mathbf{A}$. $\langle\mathbf{a}(i),\mathbf{a}(j)\rangle$ denotes the inner product of the two column vectors of $\mathbf{A}$; vec$(\mathbf{A})$ denotes vectorization of $\mathbf{A}$; $[\mathbf{A}\,|\,\mathbf{B}]$ denotes the horizontal concatenation. $\mathbf{I}_N$ and $\mathbf{0}_{N\times M}$ are the $N\times N$ identity and $N\times M$ zero matrices.

2. SYSTEM MODEL

A single link where the transmitter with a uniform linear array (ULA) of N_T antennas communicates N_S data streams to the receiver with a ULA of N_R antennas is shown in Fig. 1. At the transmitter, the precoder (consisting of an analog beamforming matrix $\mathbf{F}_{RF}\in\mathbb{C}^{N_T\times N_{RF}}$ and a digital beamforming matrix $\mathbf{F}_{BB}\in\mathbb{C}^{N_{RF}\times N_S}$) steers the N_S hybrid beams, and each hybrid beam is formed by a weighted combination (defined in the DBF) of N_{RF} steering vectors [3]. The N_{RF} steering vectors are selected from the codebook $\mathcal{F}=\{\tilde{\mathbf{f}}_{RF}(n_f), n_f=1,\cdots,N_F\}$, where $N_T\geq N_F\gg N_{RF}$. One element of $\mathcal{F}$ can be shown as [9]

The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement n^{o}619563 (MiWaveS).

$$\tilde{\mathbf{f}}_{RF}(n_f) = \frac{1}{\sqrt{N_T}}\left[1, e^{j2\pi(n_f-1)/N_T}, \cdots, e^{j2\pi(n_f-1)(N_T-1)/N_T}\right]^T. \tag{1}$$

$\mathcal{F}$ is assumed to be an orthogonal codebook[1]. The power constraint on the HBF is enforced by $\|\mathbf{F}_{RF}\mathbf{F}_{BB}\|_F^2 = N_S$.

At the receiver, the same number N_{RF} of RF devices and the same number N_S of data streams are assumed. The codebook denoted as $\mathcal{W} = \{\tilde{\mathbf{w}}_{RF}(n_w), n_w = 1, \cdots, N_W\}$ is also an orthogonal one, where $N_R \geq N_W \gg N_{RF}$ and $\tilde{\mathbf{w}}_{RF}(n_w)$ can be generated by the same rule as (1).

Via a static channel, the received signal after the combiner $\mathbf{W}_{RF}\mathbf{W}_{BB}$ at time t can be written as

$$\mathbf{r}^{(t)} = \mathbf{W}_{BB}^H\mathbf{W}_{RF}^H\mathbf{H}\mathbf{F}_{RF}\mathbf{F}_{BB}\mathbf{s}^{(t)} + \mathbf{W}_{BB}^H\mathbf{W}_{RF}^H\mathbf{z}^{(t)}, \tag{2}$$

where $\mathbf{r}^{(t)} \in \mathbb{C}^{N_S\times 1}$ is the combined received signal, $\mathbf{s}^{(t)} \in \mathbb{C}^{N_S\times 1}$ is the transmitted signal satisfying $\mathrm{E}[\mathbf{s}^{(t)}(\mathbf{s}^{(t)})^H] = \frac{1}{N_S}\mathbf{I}_{N_S}$, and $\mathbf{z}^{(t)} \in \mathbb{C}^{N_R\times 1}$ is a circularly symmetric complex Gaussian vector, $\mathbf{z}^{(t)} \sim \mathcal{CN}(\mathbf{0}_{N_R\times 1}, \sigma_z^2\mathbf{I}_{N_R})$.

mmWave channels are different to Rayleigh/Rician fading channel models, which are often assumed for centimeter wave communications. One key difference is its sparsity in spatial frequency domain [10][11]. The static channel matrix $\mathbf{H}$ in (2) can be modeled as the sum of the outer products of the array response vectors associated with P paths given by [10][12]

$$\begin{aligned}
\mathbf{H} &= \sum_{p=1}^{P}\alpha_p\mathbf{a}_A(\phi_{A,p})\mathbf{a}_D(\phi_{D,p})^H \\
&= \underbrace{\left[\mathbf{a}_A(\phi_{A,1}), \cdots, \mathbf{a}_A(\phi_{A,P})\right]}_{\mathbf{A}_A}\underbrace{\begin{bmatrix}\alpha_1 & \cdots & 0\\ \vdots & \ddots & \vdots \\ 0 & \cdots & \alpha_P\end{bmatrix}}_{\mathbf{D}}\underbrace{\begin{bmatrix}\mathbf{a}_D(\phi_{D,1})^H \\ \vdots \\ \mathbf{a}_D(\phi_{D,P})^H\end{bmatrix}}_{\mathbf{A}_D^H} \\
&= \mathbf{A}_A\mathbf{D}\mathbf{A}_D^H,
\end{aligned} \tag{3}$$

which consists of three factors: $\mathbf{A}_A \in \mathbb{C}^{N_R\times P}$ and $\mathbf{A}_D \in \mathbb{C}^{N_T\times P}$ describe the array response vectors taken from the array manifold at the receiver and transmitter respectively, and the diagonal elements of $\mathbf{D} \in \mathbb{C}^{P\times P}$ stand for the complex attenuation coefficients. Each array response vector in $\mathbf{A}_D$ can be expressed as [13]

$$\mathbf{a}_D(\phi_{D,p}) = \frac{1}{\sqrt{N_T}}\left[1, e^{j\frac{2\pi}{\lambda}\sin\phi_{D,p}\Delta_d}, \cdots, e^{j\frac{2\pi}{\lambda}\sin\phi_{D,p}(N_T-1)\Delta_d}\right]^T, \tag{4}$$

where $\phi_{D,p}, -\frac{\pi}{2} \leq \phi_{D,p} \leq \frac{\pi}{2}$, stands for the AoD for path p, λ is the wavelength at the carrier frequency, and $\Delta_d = \frac{\lambda}{2}$ is the distance between two antennas. Each array response vector in $\mathbf{A}_A$ has the similar form as (4).

3. PROPOSED LOW-COMPLEXITY BEAMFORMING METHOD BY ORTHOGONAL CODEBOOKS

The precoder and combiner design based on the SVD of $\mathbf{H}$ is usually called eigenmode transmission [3][14]. However, in practice the problem is quite intractable due to the fact that we have to estimate $\mathbf{H}$ and the computational complexity of the SVD increases exponentially with the number of antennas, given by $O(N_TN_R\cdot\min(N_T, N_R))$ [15].

To obtain the observations for the channel estimation, one can assume that the DBFs at the transmitter and receiver initially operate with $\mathbf{F}_{BB} = [\mathbf{I}_{N_S}, \mathbf{0}_{N_S\times(N_{RF}-N_S)}]^T$ and $\mathbf{W}_{BB} = [\mathbf{I}_{N_S}, \mathbf{0}_{N_S\times(N_{RF}-N_S)}]^T$ to simplify the problem. Then all the combinations of the columns of $\mathcal{F}$ and $\mathcal{W}$ are trained by using a known training sequence $\{s^{(1)}, \cdots, s^{(T)}\}$. After correlating the received signal with the training sequence, the sample mean of the observations (equivalent to the minimum variance unbiased estimator [16]) is used as the estimate for the coupling coefficient with a single combination of $\tilde{\mathbf{f}}_{RF}(n_f)$ and $\tilde{\mathbf{w}}_{RF}(n_w)$, which can be represented as

$$\begin{aligned}
&y_{n_w,n_f} \\
&= \frac{1}{T}\sum_{t=1}^{T}\frac{(s^{(t)})^*}{|s^{(t)}|^2}\left(\tilde{\mathbf{w}}_{RF}(n_w)^H\mathbf{H}\tilde{\mathbf{f}}_{RF}(n_f)\,s^{(t)} + \tilde{\mathbf{w}}_{RF}(n_w)^H\mathbf{z}^{(t)}\right) \\
&= \tilde{\mathbf{w}}_{RF}(n_w)^H\mathbf{H}\tilde{\mathbf{f}}_{RF}(n_f) + \frac{1}{T}\sum_{t=1}^{T}\frac{(s^{(t)})^*}{|s^{(t)}|^2}\tilde{\mathbf{w}}_{RF}(n_w)^H\mathbf{z}^{(t)} \\
&= \tilde{\mathbf{w}}_{RF}(n_w)^H\mathbf{H}\tilde{\mathbf{f}}_{RF}(n_f) + z'_{n_w,n_f},
\end{aligned} \tag{5}$$

where z'_{n_w,n_f} is still normally distributed but with its variance scaled by $\frac{1}{T}$ because the elements of $\mathbf{z}^{(t)}$ are independent and identically distributed (i.i.d.) Gaussian random variables and $\tilde{\mathbf{w}}_{RF}(n_w)$ contains elements with the same magnitude.

Collecting the coupling coefficients associated with all the combinations of the columns of $\mathcal{F}$ and $\mathcal{W}$ in a matrix, one obtains [7][17]

$$\mathbf{Y} = \begin{bmatrix} y_{1,1} & \cdots & y_{1,N_F} \\ \vdots & \ddots & \vdots \\ y_{N_W,1} & \cdots & y_{N_W,N_F}\end{bmatrix} = \tilde{\mathbf{W}}_{RF}^H\mathbf{H}\tilde{\mathbf{F}}_{RF} + \mathbf{Z}', \tag{6}$$

where $\tilde{\mathbf{F}}_{RF} = [\tilde{\mathbf{f}}_{RF}(1), \cdots, \tilde{\mathbf{f}}_{RF}(N_F)]$ and $\tilde{\mathbf{W}}_{RF} = [\tilde{\mathbf{w}}_{RF}(1), \cdots, \tilde{\mathbf{w}}_{RF}(N_W)]$ are the matrices consisting of all the columns of $\mathcal{F}$ and $\mathcal{W}$ respectively. Vectorizing (6) and using rules for Kronecker product, it becomes

$$\begin{aligned}
\mathrm{vec}(\mathbf{Y}) &= \underbrace{(\tilde{\mathbf{F}}_{RF}^T \otimes \tilde{\mathbf{W}}_{RF}^H)}_{\mathbf{\Phi}}\mathrm{vec}(\mathbf{H}) + \mathrm{vec}(\mathbf{Z}') \\
&= \mathbf{\Phi}\,\mathrm{vec}(\mathbf{A}_A\mathbf{D}\mathbf{A}_D^H) + \mathrm{vec}(\mathbf{Z}') \\
&= \mathbf{\Phi}(\mathbf{A}_D^* \otimes \mathbf{A}_A)\mathrm{vec}(\mathbf{D}) + \mathrm{vec}(\mathbf{Z}'),
\end{aligned} \tag{7}$$

where $\mathrm{vec}(\mathbf{Y}) \in \mathbb{C}^{N_FN_W\times 1}$ is the observation for the channel estimation. If the values of $\phi_{D,p}$ and $\phi_{A,p}$ in the array response vectors can be approximated by the values chosen from the given finite sets, e.g., $\{-90^o, -80^o, \cdots, 90^o\}$, $\mathrm{vec}(\mathbf{Y})$ can be approximated by

$$\mathrm{vec}(\mathbf{Y}) \approx \mathbf{\Phi}(\underline{\mathbf{A}}_D^* \otimes \underline{\mathbf{A}}_A)\mathrm{vec}(\underline{\mathbf{D}}) + \mathrm{vec}(\mathbf{Z}'), \tag{8}$$

where $\underline{\mathbf{A}}_D \in \mathbb{C}^{N_T\times P}$, $\underline{\mathbf{A}}_A \in \mathbb{C}^{N_R\times P}$, and the diagonal matrix $\underline{\mathbf{D}} \in \mathbb{C}^{P\times P}$ are the approximations.

3.1. Analog Beam Selection by Orthogonal Matching Pursuit

The work in [7] presents a method to obtain $\underline{\mathbf{A}}_D$ and $\underline{\mathbf{A}}_A$ which are selected from two given sets respectively. If these two sets are the

[1]The orthogonality of any two columns of $\mathcal{F}$ is defined as $\frac{\langle\tilde{\mathbf{f}}_{RF}(i), \tilde{\mathbf{f}}_{RF}(j)\rangle}{||\tilde{\mathbf{f}}_{RF}(i)||_2\cdot||\tilde{\mathbf{f}}_{RF}(j)||_2} = \begin{cases}0, & i \neq j \\ 1, & i = j\end{cases}$.

Algorithm 1: analog beam selection by OMP
Input: $\mathbf{y}_V$, $\tilde{\mathbf{F}}_{RF}$, $\tilde{\mathbf{W}}_{RF}$
Output: $\hat{\mathbf{F}}_{RF}$, $\hat{\mathbf{W}}_{RF}$
1. $\hat{\mathbf{F}}_{RF} =$ empty matrix, $\hat{\mathbf{W}}_{RF} =$ empty matrix, $\hat{\boldsymbol{\Psi}} =$ empty matrix, $\mathcal{S}_{\mathcal{F}} = \varnothing$, $\mathcal{S}_{\mathcal{W}} = \varnothing$
2. $\mathbf{y}_R = \mathbf{y}_V$
3. $\boldsymbol{\Psi} = \boldsymbol{\Phi}(\tilde{\mathbf{F}}_{RF}^* \otimes \tilde{\mathbf{W}}_{RF})$
4. for $n_{rf} = 1 : N_{RF}$
5. $\quad \mathbf{g} = \boldsymbol{\Psi}^H \mathbf{y}_R$
6. $\quad (\hat{n}_f, \hat{n}_w) = \underset{\substack{n_f \in \{1,\cdots,N_F\}\backslash \mathcal{S}_{\mathcal{F}} \\ n_w \in \{1,\cdots,N_W\}\backslash \mathcal{S}_{\mathcal{W}}}}{\arg\max} |g_i|$
   $\quad$ where $i = (n_f - 1)N_W + n_w$
7. $\quad \hat{\mathbf{F}}_{RF} = [\hat{\mathbf{F}}_{RF} \,|\, \tilde{\mathbf{f}}_{RF}(\hat{n}_f)]$
   $\quad \hat{\mathbf{W}}_{RF} = [\hat{\mathbf{W}}_{RF} \,|\, \tilde{\mathbf{w}}_{RF}(\hat{n}_w)]$
8. $\quad \hat{\boldsymbol{\Psi}} = [\hat{\boldsymbol{\Psi}} \,|\, \boldsymbol{\Phi}(\tilde{\mathbf{f}}_{RF}(\hat{n}_f)^* \otimes \tilde{\mathbf{w}}_{RF}(\hat{n}_w))]$
9. $\quad \mathbf{y}_R = (\mathbf{I}_{N_F N_W} - \hat{\boldsymbol{\Psi}}(\hat{\boldsymbol{\Psi}}^H \hat{\boldsymbol{\Psi}})^{-1} \hat{\boldsymbol{\Psi}}^H)\mathbf{y}_V$
10. $\quad \mathcal{S}_{\mathcal{F}} = \mathcal{S}_{\mathcal{F}} \cup \{\hat{n}_f\}$ and $\mathcal{S}_{\mathcal{W}} = \mathcal{S}_{\mathcal{W}} \cup \{\hat{n}_w\}$
11. end

same as $\mathcal{F}$ and $\mathcal{W}$, (8) can be further shown as (see the Appendix)

$$\text{vec}(\mathbf{Y}) \approx \boldsymbol{\Phi}(\mathbf{F}_{RF}^* \otimes \mathbf{W}_{RF})\text{vec}(\underline{\mathbf{D}}(1:N_{RF}, 1:N_{RF})) + \text{vec}(\mathbf{Z}'). \tag{9}$$

Express (9) in a simpler form, one has

$$\mathbf{y}_V \approx \boldsymbol{\Phi}(\mathbf{F}_{RF}^* \otimes \mathbf{W}_{RF})\underline{\mathbf{d}}_V + \mathbf{z}'_V, \tag{10}$$

where $\mathbf{y}_V = \text{vec}(\mathbf{Y})$, $\underline{\mathbf{d}}_V = \text{vec}(\underline{\mathbf{D}}(1:N_{RF}, 1:N_{RF}))$, and $\mathbf{z}'_V = \text{vec}(\mathbf{Z}')$.

In (10), the design criterion used for the ABFs is to minimize the 2-norm of the error between the observation and the reconstructed signal

$$(\hat{\mathbf{F}}_{RF}, \hat{\mathbf{W}}_{RF}) = \underset{\mathbf{F}_{RF}, \mathbf{W}_{RF}}{\arg\min} \|\mathbf{y}_V - \boldsymbol{\Phi}(\mathbf{F}_{RF}^* \otimes \mathbf{W}_{RF})\underline{\mathbf{d}}_V\|_2 ,$$
$$\text{s.t.} \begin{cases} \mathbf{f}_{RF}(n_{rf}) \in \mathcal{F}, \mathbf{w}_{RF}(n_{rf}) \in \mathcal{W}, n_{rf} = 1, \cdots, N_{RF}, \\ \|\underline{\mathbf{d}}_V\|_0 = N_{RF}, \\ \text{rank}(\mathbf{F}_{RF}) = N_{RF}, \text{rank}(\mathbf{W}_{RF}) = N_{RF}. \end{cases} \tag{11}$$

where $\underline{\mathbf{d}}_V$ is an N_{RF}-sparse signal, and the full rank constraints ensure that there are no repeated column vectors in both $\hat{\mathbf{F}}_{RF}$ and $\hat{\mathbf{W}}_{RF}$. (11) can be solved by the OMP, which is a greedy algorithm providing a solution to find a column that is most correlated with the current residual in every iteration (see **Algorithm 1**) [6].

3.2. Simplified Analog Beam Selection by Orthogonal Codebooks

In **Algorithm 1**, if $\mathcal{F}$ and $\mathcal{W}$ are orthogonal codebooks, $\boldsymbol{\Psi}$ is essentially the identity matrix,

$$\begin{aligned} \boldsymbol{\Psi} &= \boldsymbol{\Phi}(\tilde{\mathbf{F}}_{RF}^* \otimes \tilde{\mathbf{W}}_{RF}) \\ &= \underbrace{(\tilde{\mathbf{F}}_{RF}^T \otimes \tilde{\mathbf{W}}_{RF}^H)}_{\boldsymbol{\Phi} \text{ in (7)}}(\tilde{\mathbf{F}}_{RF}^* \otimes \tilde{\mathbf{W}}_{RF}) \\ &= \tilde{\mathbf{F}}_{RF}^T \tilde{\mathbf{F}}_{RF}^* \otimes \tilde{\mathbf{W}}_{RF}^H \tilde{\mathbf{W}}_{RF} \\ &= \mathbf{I}_{N_F N_W}. \end{aligned} \tag{12}$$

Therefore, $\{\boldsymbol{\psi}(i), i = 1, \cdots, N_F N_W\}$ is the standard basis for $\mathbb{C}^{N_F N_W \times 1}$, where $\boldsymbol{\psi}(i)$ is the i^{th} column of $\boldsymbol{\Psi}$, and the i^{th} element of $\mathbf{g}$ (i.e., the inner product of $\boldsymbol{\psi}(i)$ and the residual $\mathbf{y}_R$ at the $(n_{rf}+1)^{\text{th}}$ iteration in **Algorithm 1** Step 6) can be shown as

$$\begin{aligned} g_i &= \boldsymbol{\psi}(i)^H \mathbf{y}_R \\ &\overset{(a)}{=} \boldsymbol{\psi}(i)^H(\mathbf{I}_{N_F N_W} - \hat{\boldsymbol{\Psi}}(\underbrace{\hat{\boldsymbol{\Psi}}^H \hat{\boldsymbol{\Psi}}}_{\mathbf{I}_{n_{rf}}})^{-1}\hat{\boldsymbol{\Psi}}^H)\mathbf{y}_V \\ &= \boldsymbol{\psi}(i)^H(\mathbf{I}_{N_F N_W} - \hat{\boldsymbol{\Psi}}\hat{\boldsymbol{\Psi}}^H)\mathbf{y}_V \\ &= \boldsymbol{\psi}(i)^H\mathbf{y}_V - \boldsymbol{\psi}(i)^H(\hat{\boldsymbol{\psi}}(1)\hat{\boldsymbol{\psi}}(1)^H + \cdots + \hat{\boldsymbol{\psi}}(n_{rf})\hat{\boldsymbol{\psi}}(n_{rf})^H)\mathbf{y}_V \\ &= \begin{cases} 0, & \boldsymbol{\psi}(i) \in \{\hat{\boldsymbol{\psi}}(1), \cdots, \hat{\boldsymbol{\psi}}(n_{rf})\} \\ \boldsymbol{\psi}(i)^H\mathbf{y}_V, & \boldsymbol{\psi}(i) \notin \{\hat{\boldsymbol{\psi}}(1), \cdots, \hat{\boldsymbol{\psi}}(n_{rf})\}. \end{cases} \end{aligned} \tag{13}$$

where (a) uses the relationship between $\mathbf{y}_R$ and $\mathbf{y}_V$ in **Algorithm 1** Step 9. The full rank constraints in (11) ensure that the first condition $\boldsymbol{\psi}(i) \in \{\hat{\boldsymbol{\psi}}(1), \cdots, \hat{\boldsymbol{\psi}}(n_{rf})\}$ in (13) will not happen. On the other hand, from the second condition $\boldsymbol{\psi}(i) \notin \{\hat{\boldsymbol{\psi}}(1), \cdots, \hat{\boldsymbol{\psi}}(n_{rf})\}$, we can find that

$$g_i = \boldsymbol{\psi}(i)^H \mathbf{y}_V \overset{(b)}{=} y_{n_w, n_f}, \tag{14}$$

where (b) follows from the fact that $\boldsymbol{\psi}(i)$ is a standard basis vector so that $\boldsymbol{\psi}(i)^H\mathbf{y}_V$ returns the i^{th} entry of $\mathbf{y}_V$, which is also the $(n_w, n_f)^{\text{th}}$ entry of $\mathbf{Y}$. Consequently, it is not necessary to update the residual $\mathbf{y}_R$ in Step 9 in every iteration; instead, the analog beams can be decided according to the magnitude of the codebook training results $\{y_{n_w,n_f}, n_w = 1, \cdots, N_W, n_f = 1, \cdots, N_F\}$ obtained from (5). **Algorithm 1** therefore can be rewritten as

Revised Algorithm 1: proposed simplified analog beam selection (it only shows the revised for loop)
for $n_{rf} = 1 : N_{RF}$
$\quad (\hat{n}_f, \hat{n}_w) = \underset{\substack{n_f \in \{1,\cdots,N_F\}\backslash \mathcal{S}_{\mathcal{F}} \\ n_w \in \{1,\cdots,N_W\}\backslash \mathcal{S}_{\mathcal{W}}}}{\arg\max} \left|y_{n_w,n_f}\right|$
$\quad \hat{\mathbf{F}}_{RF} = [\hat{\mathbf{F}}_{RF} \,|\, \tilde{\mathbf{f}}_{RF}(\hat{n}_f)]$ and $\hat{\mathbf{W}}_{RF} = [\hat{\mathbf{W}}_{RF} \,|\, \tilde{\mathbf{w}}_{RF}(\hat{n}_w)]$
$\quad \mathcal{S}_{\mathcal{F}} = \mathcal{S}_{\mathcal{F}} \cup \{\hat{n}_f\}$ and $\mathcal{S}_{\mathcal{W}} = \mathcal{S}_{\mathcal{W}} \cup \{\hat{n}_w\}$
end

An alternative explanation for **Revised Algorithm 1** is that $\tilde{\mathbf{W}}_{RF}^H$ and $\tilde{\mathbf{F}}_{RF}$ in (6) can be viewed as the DFT and IDFT matrices respectively. Consequently, $\{y_{n_w,n_f}, n_w = 1, \cdots, N_W, n_f = 1, \cdots, N_F\}$ are in essence the *sparse* spatial frequency domain signals, and their magnitude leads to the desired analog beams with equal weighting defined in $\mathbf{F}_{BB}$ and $\mathbf{W}_{BB}$.

3.3. DBF Design by Low-Dimensional SVD

Based on the selected analog beams from **Revised Algorithm 1**, one has the effective channel matrix of size $N_{RF} \times N_{RF}$

$$\mathbf{H}_E = \hat{\mathbf{W}}_{RF}^H \mathbf{H} \hat{\mathbf{F}}_{RF} \overset{\text{SVD}}{=} \mathbf{U}_E \boldsymbol{\Sigma}_E \mathbf{V}_E^H, \tag{15}$$

where the columns of $\mathbf{V}_E \in \mathbb{C}^{N_{RF} \times N_{RF}}$ and $\mathbf{U}_E \in \mathbb{C}^{N_{RF} \times N_{RF}}$ are, respectively, the right and left singular vectors of $\mathbf{H}_E$, and the diagonal elements of $\boldsymbol{\Sigma}_E \in \mathbb{R}^{N_{RF} \times N_{RF}}$ are the singular values of $\mathbf{H}_E$. Let $\hat{\mathbf{F}}_{BB} = \mathbf{V}_E(1:N_S)$ and $\hat{\mathbf{W}}_{BB} = \mathbf{U}_E(1:N_S)$, then the transmit power can be optimally allocated to the N_S data streams to maximize the system throughput [14]. Although the SVD of $\mathbf{H}_E$ is needed for the DBF reconstruction, the computational complexity is much lower than that of $\mathbf{H} \in \mathbb{C}^{N_R \times N_T}$, which is required for the HBF reconstruction presented in [3].

4. SIMULATION RESULTS

The simulation parameters are listed as follows. $N_T = N_R = N_F = N_W = 32$; $N_{RF} = N_S = 2$ or 4; the codebook training time $T = 512$; $\sigma_z^2 = \frac{1}{N_S} 10^{-\gamma/10}$, where γ (dB) is the given SNR in the simulations. Clustered-based channel models are simulated; specifically, the channel model with P paths in (3) can be rewritten as the sum of L clusters with M rays ($P = LM$)

$$\begin{aligned} \mathbf{H} &= \sum_{p=1}^{P} \alpha_p \mathbf{a}_A(\phi_{A,p}) \mathbf{a}_D(\phi_{D,p})^H \\ &= \sum_{l=1}^{L} \underbrace{\sum_{m=1}^{M} \alpha_{l,m} \mathbf{a}_A(\phi_{A,l,m}) \mathbf{a}_D(\phi_{D,l,m})^H}_{\text{cluster } l}, \end{aligned} \tag{16}$$

where $\mathbf{a}_A(\phi_{A,l,m})$ and $\mathbf{a}_D(\phi_{D,l,m})$ have the similar forms as (4). $\alpha_{l,m} = \beta_l e^{j\varphi_{l,m}}$, where $\beta_l \in \mathbb{R}$ stands for the path loss for cluster l and the initial phase $\varphi_{l,m} \sim \mathcal{U}(0, 2\pi)$ is different for each ray in cluster l. Two simulated channel models are detailed below:

1. One dominant light-of-sight (LoS) cluster ($L = 1, M = 8, P = 8$) with the path loss β_l of $126.2 - 12\log_{10}(d) + 4.4x$ dB, where $d = 50$ m is the distance between the transmitter and receiver and $x \sim \mathcal{N}(0, 1)$ [11]. LoS AoDs are generated by $\phi_{D,l,m} = \overline{\phi}_l + \Delta\phi_{\text{LoS},l,m}$, where the same $\overline{\phi}_l \sim \mathcal{U}(-\frac{\pi}{2}, \frac{\pi}{2})$ is applied to all the rays in the cluster and the angular spread in the cluster is given by $\Delta\phi_{\text{LoS},l,m} = 7^\circ + 4.2^\circ x$ [11]. The same procedure is used to generate LoS AoAs.
2. One LoS and additional three NLoS clusters ($L = 4, M = 8, P = 32$). The NLoS path attenuation is given by $43.2 + 49\log_{10}(d) + 10.3x$ dB. NLoS AoDs are generated by $\phi_{D,l,m} = \overline{\phi}_l + \Delta\phi_{\text{NLoS},l,m}$, where $\Delta\phi_{\text{NLoS},l,m} = 3.7^\circ + 2.3^\circ x$ [11]. The same procedure is used to generate NLoS AoAs.

Assume Gaussian signaling, given $\mathbf{F}_{BB}$, $\mathbf{F}_{RF}$, $\mathbf{W}_{BB}$, $\mathbf{W}_{RF}$, the system throughput is shown as [18]

$$\begin{aligned} R = \log_2 \Big| \mathbf{I}_{N_S} + \frac{1}{N_S} \mathbf{R}_z^{-1} \left(\mathbf{W}_{BB}^H \mathbf{W}_{RF}^H \mathbf{H} \mathbf{F}_{RF} \mathbf{F}_{BB} \right) \cdot \\ \left(\mathbf{W}_{BB}^H \mathbf{W}_{RF}^H \mathbf{H} \mathbf{F}_{RF} \mathbf{F}_{BB} \right)^H \Big|, \end{aligned} \tag{17}$$

where $\mathbf{R}_z = \sigma_z^2 \mathbf{W}_{BB}^H \mathbf{W}_{RF}^H \mathbf{W}_{RF} \mathbf{W}_{BB}$ is the noise covariance matrix after combining.

Fig. 2 shows the achievable data rates by the reference method [3] and the proposed one. In [3], $\mathbf{F}_{BB}$, $\mathbf{F}_{RF}$, $\mathbf{W}_{BB}$, and $\mathbf{W}_{RF}$ are reconstructed based on the right and left singular vectors of the perfect channel matrix. In the proposed method, the columns of $\mathbf{F}_{RF}$ and $\mathbf{W}_{RF}$ are selected according to the magnitude of the codebook training results $\{y_{n_w,n_f}, n_w = 1, \cdots, N_W, n_f = 1, \cdots, N_F\}$ (see **Revised Algorithm 1**). Without considering power allocation at the transmitter, such as water filling [14], the digital beamforming matrices can be assumed to be the identity matrices because the achievable data rates are the same as the result by (15) when $N_{RF} = N_S$. Thus, the simulation results of the proposed scheme are obtained by the assumptions that $\hat{\mathbf{F}}_{BB} = \mathbf{I}_{N_S}$ and $\hat{\mathbf{W}}_{BB} = \mathbf{I}_{N_S}$.

From Fig. 2 we can find that the resulting data rates of the proposed method are almost the same as the reference. Although the proposed one is limited to the orthogonal codebooks, the joint channel estimation and beamforming problem becomes tractable because only the codebook training procedure is required.

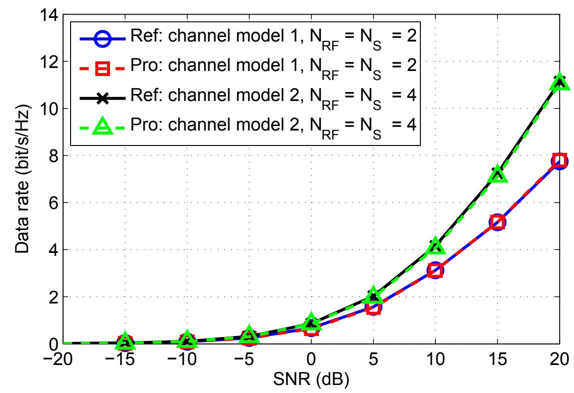

Fig. 2. The achievable data rates by the reference and the proposed methods in the two simulated environments. The total power of all the paths is normalized to one.

5. CONCLUSION

This paper presents a low-complexity beamforming method. It shows that given the orthogonal codebooks the analog beams can be selected according to the magnitude of the codebook training results, and the achievable data rates are nearly the same as the one based on the perfect CSI. The key idea is to exploit the *implicit* knowledge of the channel to rapidly reconstruct ABFs at the transmitter and receiver, and then the DBFs can be implemented by the low-dimensional SVD of the *explicitly* estimated effective channel matrix.

6. APPENDIX

In [7], given the sets consisting of all the candidates of the columns of $\underline{\mathbf{A}}_D$ and $\underline{\mathbf{A}}_A$ in (8), if the selected array response matrices $\hat{\mathbf{A}}_D$ and $\hat{\mathbf{A}}_A$ satisfy $\hat{\mathbf{A}}_D^H \hat{\mathbf{A}}_D = \mathbf{I}_P$ and $\hat{\mathbf{A}}_A^H \hat{\mathbf{A}}_A = \mathbf{I}_P$, the HBFs can be reconstructed directly based on $\hat{\mathbf{A}}_D$ and $\hat{\mathbf{A}}_A$ without SVD. Then the design criterion that can be used for the precoder reconstruction is to minimize the Frobenius norm of the error between $\hat{\mathbf{A}}_D(1:N_S)$ and the reconstructed precoder of size $N_T \times N_S$,

$$\begin{aligned} (\hat{\mathbf{F}}_{RF}, \hat{\mathbf{F}}_{BB}) &= \arg \min_{\mathbf{F}_{RF}, \mathbf{F}_{BB}} \left\| \hat{\mathbf{A}}_D(1:N_S) - \mathbf{F}_{RF}\mathbf{F}_{BB} \right\|_F, \\ \text{s.t.} &\begin{cases} \mathbf{f}_{RF}(n_{rf}) \in \mathcal{F}, n_{rf} = 1, \cdots, N_{RF}, \\ \|\mathbf{F}_{RF}\mathbf{F}_{BB}\|_F^2 = N_S. \end{cases} \end{aligned} \tag{18}$$

If the same set $\mathcal{F}$ is used to obtain $\underline{\mathbf{A}}_D$ and $\mathbf{F}_{RF}$, the optimal solution to $\mathbf{F}_{RF}$ based on $\hat{\mathbf{A}}_D$ in (18) is $\hat{\mathbf{F}}_{RF} = \hat{\mathbf{A}}_D(1 : N_{RF})$, and therefore the optimal solution to $\mathbf{F}_{BB}$ is $\hat{\mathbf{F}}_{BB} = [\mathbf{I}_{N_S}, \mathbf{0}_{N_S \times (N_{RF}-N_S)}]^T$. Similarly, if the same set $\mathcal{W}$ is used to obtain $\underline{\mathbf{A}}_A$ and $\mathbf{W}_{RF}$, we have $\hat{\mathbf{W}}_{RF} = \hat{\mathbf{A}}_A(1 : N_{RF})$ and $\hat{\mathbf{W}}_{BB} = [\mathbf{I}_{N_S}, \mathbf{0}_{N_S \times (N_{RF}-N_S)}]^T$. As a result, the first N_{RF} columns of $\underline{\mathbf{A}}_D$ and $\underline{\mathbf{A}}_A$ (which are associated with the N_{RF} ($N_{RF} \le P$) strongest paths) in (8) can be regarded as the analog beamforming matrices and replaced by $\mathbf{F}_{RF}$ and $\mathbf{W}_{RF}$.

7. REFERENCES

[1] T. S. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. N. Wong, J. K. Schulz, M. Samimi, and F. Gutierrez, "Millimeter wave mobile communications for 5G cellular: It will work!," *IEEE Access*, vol. 1, pp. 335–349, 2013.

[2] T. S. Rappaport, R. C. Daniels, and J. N. Murdock, *Millimeter Wave Wireless Communications*, Prentice Hall, 2014.

[3] O. E. Ayach, R. W. Heath, S. Abu-Surra, S. Rajagopal, and Z. Pi, "Low complexity precoding for large millimeter wave mimo systems," in *2012 IEEE International Conference on Communications (ICC)*, June 2012, pp. 3724–3729.

[4] A. Alkhateeb, O. El 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.

[5] S. Han, C. l. I, Z. Xu, and C. Rowell, "Large-scale antenna systems with hybrid analog and digital beamforming for millimeter wave 5G," *IEEE Communications Magazine*, vol. 53, no. 1, pp. 186–194, January 2015.

[6] T. 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, July 2011.

[7] H. L. Chiang, T. Kadur, W. Rave, and G. Fettweis, "Low-complexity spatial channel estimation and hybrid beamforming for millimeter wave links," in *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, 2016, pp. 942–948.

[8] S. Foucart and H. Rauhut, *A Mathematical Introduction to Compressive Sensing*, Birkhäuser Basel, 2013.

[9] H. L. Chiang, T. Kadur, and G. Fettweis, "Analyses of orthogonal and non-orthogonal steering vectors at millimeter wave systems," in *2016 IEEE 17th International Symposium on A World of Wireless, Mobile and Multimedia Networks (WoWMoM)*, June 2016, pp. 1–6.

[10] T. A. Thomas, H. C. Nguyen, G. R. MacCartney, and T. S. Rappaport, "3D mmwave channel model proposal," in *2014 IEEE 80th Vehicular Technology Conference (VTC2014-Fall)*, Sept 2014, pp. 1–6.

[11] T. S. Rappaport, G. R. MacCartney, M. K. Samimi, and S. Sun, "Wideband millimeter-wave propagation measurements and channel models for future wireless communication system design," *IEEE Transactions on Communications*, vol. 63, no. 9, pp. 3029–3056, Sept 2015.

[12] P. Kyösti et al., "Winner II channel models," Tech. Rep., IST-4-027756 WINNER II D1.1.2 v 1.2, 2007.

[13] J. C. Liberti and T. S. Rappaport, *Smart antennas for wireless communications: IS-95 and third generation CDMA applications*, Prentice Hall, 1999.

[14] E. Telatar, "Capacity of multi-antenna gaussian channels," *European Transactions on Telecommunications*, vol. 10, pp. 585–595, 1999.

[15] Gene H. Golub and Charles F. Van Loan, *Matrix Computations (3rd Ed.)*, Johns Hopkins University Press, 1996.

[16] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, Prentice Hall, 1997.

[17] R. Méndez-Rial, C. Rusu, A. Alkhateeb, N. González-Prelcic, and R. W. Heath, "Channel estimation and hybrid combining for mmwave: Phase shifters or switches?," in *2015 Information Theory and Applications Workshop (ITA)*, Feb 2015, pp. 90–97.

[18] A. Goldsmith, S. A. Jafar, N. Jindal, and S. Vishwanath, "Capacity limits of mimo channels," *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 5, pp. 684–702, June 2003.