2016 IEEE 27th Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communications - (PIMRC): Fundamentals and PHY

Low-Complexity Spatial Channel Estimation and Hybrid Beamforming for Millimeter Wave Links

Hsiao-Lan Chiang, Tobias Kadur, Wolfgang Rave, and Gerhard Fettweis
Technische Universität Dresden, Email: {hsiao-lan.chiang, tobias.kadur, rave, gerhard.fettweis}@ifn.et.tu-dresden.de

***Abstract*—Efficiently estimating spatial channel properties, such as angles of arrival and departure, for a hybrid beamforming (HBF) architecture is one of the crucial challenges to overcome at millimeter wave (mmW) systems. To this end, we propose an algorithm variant to a recently proposed approach [8] based on ideas borrowed from the compressed sensing literature and l_0-norm minimization, which exploit the fact that the number of significant channel echoes is rather small for limited beamwidth. Our modified algorithm eliminates high dimensional singular value decomposition (SVD) of the estimated channel matrix in the original method by employing the orthogonality between the selected array propagation vectors. It is demonstrated that HBF design *without* SVD of the estimated channel matrix can achieve essentially the same capacity as the one *with* SVD. Moreover, the feedback overhead required for the beamforming systems can be significantly reduced by the proposed method.**

I. Introduction

With the rapid increase of data rate in wireless communication, bandwidth shortage is getting more critical. Therefore, there is a growing interest in using millimeter wave (mmW) for future wireless communications taking advantage of the enormous amount of available spectrum [1]. Measurements of most large and small scale parameters for mmW channels in urban areas at 73 GHz had been presented in [2]. It seems that path loss in such environment is very severe, and in order to improve capacity and service quality, mmW together with massive multiple input multiple output is a promising approach [3].

Most of today's systems operating in lower carrier frequency bands ($\leq$ 6 GHz) are based on *digital* beamforming (DBF) architectures. In this architecture, each antenna has its own RF chain, including digital-to-analog converter (DAC) at the transmitter and analog-to-digital converter (ADC) at the receiver. However, to exploit the huge available bandwidth in mmW bands, ADCs and DACs have to run at several Giga samples per second. This makes it infeasible to equip each antenna with its own RF chain due to high implementation cost and power consumption. Therefore, a combination of *analog* beamforming (ABF) [4][5] (operating in passband) and DBF (operating in baseband) will be implemented to achieve higher data rate with acceptable complexity. We denote this combination as *hybrid beamforming* (HBF) [6][7]. ABF is typically implemented at each antenna using a finite set of possible phase shifts, where one entry can be selected from a codebook to produce a beam pattern with a specific main lobe direction. Selecting several beam patterns simultaneously provides, after down conversion, the input to DBF that is to be optimally combined according to some criteria, e.g. to maximize mutual information.

The problem of joint spatial channel estimation (CE, it is assumed to be flat in frequency domain) and HBF is complicated due to the fact that ABF is not as flexible as DBF because equal magnitude of the analog beam pattern gain is used for all antenna signals. To obtain full channel information, the coupling of ABFs across the channel is needed and therefore exhaustive analog beam training is required in the beginning [8]. Typically, the problem of both CE and HBF can be formulated as an l_0-norm optimization problem [9] since mmW channels consist of a quite small number of significant paths, meaning that they are sparse in angular (angles of arrival and departure (AoAs, AoDs)) or the equivalent spatial frequency domain. The solution to the problem proposed in [6][8] works as follows. First, the channel is estimated by orthogonal matching pursuit (OMP) [10] by observations based on all combinations between the analog beamformers at the transmitter and the receiver. Then, to complete the HBF design at the transmitter (precoder reconstruction) and the receiver (combiner reconstruction), again OMP is employed based on the estimated channel matrix [6]. This solution achieves promising data rate. However, its computational complexity places a huge burden at the receiver. Another problem is the feedback overhead. If the precoder reconstruction is implemented at the transmitter, the right singular vectors of the estimated channel have to be sent to the transmitter. Or if it is implemented at the receiver, then the codebook indices of the analog beamformers and the matrix of DBF have to be sent to the transmitter. Both need quite a lot feedback overhead.

This paper presents a novel method to reduce the computational complexity and the feedback overhead. By orthogonality of the selected array propagation vectors in the proposed CE, the precoder and the combiner can be reconstructed *without* singular value decomposition (SVD) of the estimated channel matrix and there is no data rate loss. Also, regarding the feedback overhead, only the codebook indices of the selected array propagation vectors have to be sent to the transmitter. Compared to the reference methods, it shows that the proposed one can achieve the desirable data rate with the least feedback overhead.

This paper is organized as follows: Section II describes the system and the mmW channel models. Section III elaborates the problem and the proposed solution to low-complexity joint CE and HBF designs. Simulation results are presented in

978-1-5090-3254-9/16/$31.00 ©2016 European Union

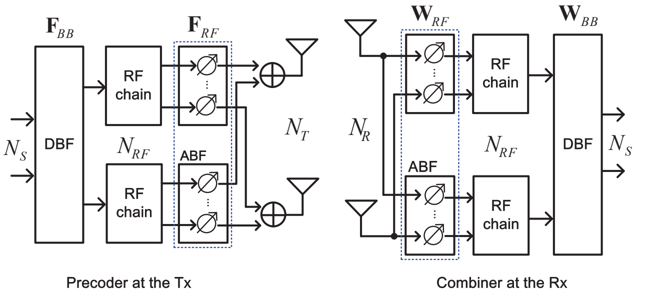

Figure 1: The beamforming system diagram.

Section IV, and we conclude our work in Section V.

We use the following notation throughout this paper. a is a scalar, $\mathbf{a}$ is a vector, and $\mathbf{A}$ is a matrix. a_i is the i^{th} entry of $\mathbf{a}$; $a_{i,j}$ or $(\mathbf{A})_{i,j}$ is the $(i,j)^{\text{th}}$ entry of $\mathbf{A}$; $\mathbf{a}(i)$ is the i^{th} column vector of $\mathbf{A}$; and $\mathbf{A}(1:N)$ denotes the first N column vectors $[\mathbf{a}(1),\cdots,\mathbf{a}(N)]$ of $\mathbf{A}$. $\mathbf{A}^*$, $\mathbf{A}^H$, and $\mathbf{A}^T$ denote complex conjugate, Hermitian transpose and transpose of $\mathbf{A}$, respectively. $\|\mathbf{a}\|_0$ is the l_0-norm of $\mathbf{a}$ [11]; $\|\mathbf{a}\|_2$ is the 2-norm of $\mathbf{a}$; $|\mathbf{A}|$ and $\|\mathbf{A}\|_F$ are the determinant and Frobenius norm of $\mathbf{A}$, respectively. $\text{diag}(\mathbf{A})$ is the vector formed by the diagonal elements of $\mathbf{A}$; $\text{vec}(\mathbf{A})$ is vectorization of $\mathbf{A}$; $\text{rank}(\mathbf{A})$ is the (column) rank of $\mathbf{A}$; and $[\mathbf{A}\,|\,\mathbf{B}]$ denotes horizontal concatenation. $\mathbf{I}_N$ is the $N\times N$ identity matrix, and $\mathbf{0}_{N\times 1}$ is the $N\times 1$ zero vector.

II. System Models

In the assumed heterogeneous system [12], uplink and downlink transmission timing between a transmitter and a receiver is synchronized. Fig. 1 shows a single link where the transmitter with N_T antennas communicates N_S data streams to the receiver with N_R antennas. The transmitter is equipped with a precoder, which consists of an $N_T\times N_{RF}$ ABF matrix $\mathbf{F}_{RF}$ and an $N_{RF}\times N_S$ DBF matrix $\mathbf{F}_{BB}$ [6]-[8]. The precoder steers the N_S data beams supported by HBF, where each hybrid beam is formed by a weighted combination (defined in the DBF) of the N_{RF} analog beams. The N_{RF} analog beams at the transmitter are selected from a codebook defined as a matrix $\tilde{\mathbf{F}}_{RF}$ of size $N_T\times N_F$, shown as

$$\tilde{\mathbf{F}}_{RF}=\frac{1}{\sqrt{N_T}}\begin{bmatrix}1 & 1 & \cdots & 1\\ 1 & e^{j2\pi\frac{1}{N_F}} & \cdots & e^{j2\pi\frac{N_F-1}{N_F}}\\ \vdots & \vdots & \ddots & \vdots\\ 1 & e^{j2\pi\frac{N_T-1}{N_F}} & \cdots & e^{j2\pi\frac{(N_F-1)(N_T-1)}{N_F}}\end{bmatrix}. \tag{1}$$

We assume that $N_F\geq N_{RF}$ (or even $\gg N_{RF}$). Considering the codebook size, when $N_F=N_T$, the matrix $\tilde{\mathbf{F}}_{RF}$ can be chosen as a IDFT matrix. The power constraint on each hybrid beam is enforced by $\|\mathbf{F}_{RF}\mathbf{f}_{BB}(n_s)\|_2^2=1, n_s=1,\cdots,N_S$, where $\mathbf{f}_{BB}(n_s)$ is the n_s^{th} column of $\mathbf{F}_{BB}$. The codebook $\tilde{\mathbf{W}}_{RF}\in\mathbb{C}^{N_R\times N_W}$ used for the ABF at the receiver has the similar structure as (1). The same number N_{RF} of RF devices and the same number N_S of data streams are assumed at the receiver.

The received signal after the combiner $\mathbf{W}_{RF}\mathbf{W}_{BB}$ can be written as

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

where $\mathbf{r}$ is the $N_S\times 1$ combined received data vector, $\mathbf{s}$ is the transmitted data vector satisfying $\text{E}[\mathbf{s}\mathbf{s}^H]=\frac{1}{N_S}\mathbf{I}_{N_S}$, and $\mathbf{z}$ is the $N_R\times 1$ additive white Gaussian noise (AWGN) vector with the complex Gaussian random variable $z_{n_r}\sim\mathcal{CN}(0,\sigma_z^2)$, $n_r=1,\cdots,N_R$.

mmW channel 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 [2]. The channel matrix $\mathbf{H}$ in (2) can be modelled as the sum of outer products of array propagation vectors associated with N_P paths,

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

which consists of three factors: the matrix $\mathbf{A}_A\in\mathbb{C}^{N_R\times N_P}$ with the array propagation vectors taken from the array manifold at the receiver, the path loss matrix $\mathbf{D}\in\mathbb{C}^{N_P\times N_P}$ with the complex attenuation coefficient α_p for each path p, and the matrix $\mathbf{A}_D\in\mathbb{C}^{N_T\times N_P}$ describing the array propagation vectors from the array manifold at the transmitter. Each array propagation vector in $\mathbf{A}_D$ can be expressed as [5]

$$\mathbf{a}_D(p)=\frac{1}{\sqrt{N_T}}\left[1,e^{j2\pi\lambda_0^{-1}\sin\phi_p\Delta_d},\cdots,e^{j2\pi\lambda_0^{-1}\sin\phi_p(N_T-1)\Delta_d}\right]^T, \tag{4}$$

where ϕ_p stands for the p^{th} AoD, $\Delta_d\triangleq\frac{\lambda_0}{2}$ is the distance between two antennas, and λ_0 is the wavelength at the carrier frequency. Each array propagation vector in $\mathbf{A}_A$ has the similar structure as (4). Assume Gaussian signaling [13], given $\mathbf{F}_{BB}$, $\mathbf{F}_{RF}$, $\mathbf{W}_{RF}$, $\mathbf{W}_{BB}$, the system throughput is shown as

$$\begin{aligned}C=\log_2\Big|\mathbf{I}_{N_S}+\frac{1}{N_S}\cdot\mathbf{R}_z^{-1}\cdot\left(\mathbf{W}_{BB}^H\mathbf{W}_{RF}^H\mathbf{H}\mathbf{F}_{RF}\mathbf{F}_{BB}\right)\cdot\\ \left(\mathbf{W}_{BB}^H\mathbf{W}_{RF}^H\mathbf{H}\mathbf{F}_{RF}\mathbf{F}_{BB}\right)^H\Big|,\end{aligned} \tag{5}$$

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.

III. Joint CE and HBF with Limited Feedback

To simplify the CE problem, let us assume that the DBFs at the transmitter and the receiver initially operates with $\mathbf{F}_{BB}=\mathbf{I}$ and $\mathbf{W}_{BB}=\mathbf{I}$. To obtain the observations for CE, all the combinations of the vectors defined in the codebooks $\tilde{\mathbf{W}}_{RF}$ and $\tilde{\mathbf{F}}_{RF}$ are trained [8] by using a known training sequence $\{s_1,\ldots,s_T\}$ of length T (the coherence time of the channel is assumed to be sufficiently long). After removing the training signals, the average of the noisy observations for the coupling coefficient with a single beamformer combination of $\tilde{\mathbf{f}}_{RF}(l)$ and $\tilde{\mathbf{w}}_{RF}(k)$ can be written as

$$\begin{aligned} y_{k,l} &= \frac{1}{T}\sum_{t=1}^{T}\frac{1}{s_t}\left(\tilde{\mathbf{w}}_{RF}^H(k)\mathbf{H}\tilde{\mathbf{f}}_{RF}(l)\,s_t + \tilde{\mathbf{w}}_{RF}^H(k)\mathbf{z}(t)\right) \\ &= \tilde{\mathbf{w}}_{RF}^H(k)\mathbf{H}\tilde{\mathbf{f}}_{RF}(l) + \frac{1}{T}\sum_{t=1}^{T}\frac{1}{s_t}\tilde{\mathbf{w}}_{RF}^H(k)\mathbf{z}(t) \\ &= \tilde{\mathbf{w}}_{RF}^H(k)\mathbf{H}\tilde{\mathbf{f}}_{RF}(l) + z'_k, \\ k &= 1,\cdots,N_W, l = 1,\cdots,N_F, \end{aligned} \tag{6}$$

where $\tilde{\mathbf{w}}_{RF}(k)$ is the k^{th} steering vector of $\tilde{\mathbf{W}}_{RF}$, $\tilde{\mathbf{f}}_{RF}(l)$ is the l^{th} steering vector of $\tilde{\mathbf{F}}_{RF}$, and z'_k is the averaged noise coupling with $\tilde{\mathbf{w}}_{RF}(k)$. Collecting the coupling coefficients associated with all the combinations of $\tilde{\mathbf{w}}_{RF}(k)$ and $\tilde{\mathbf{f}}_{RF}(l)$ in a matrix, we obtain

$$\begin{aligned} \mathbf{Y} &= \begin{bmatrix} y_{1,1} & \cdots & y_{1,N_F} \\ \vdots & \ddots & \vdots \\ y_{N_W,1} & \cdots & y_{N_W,N_F} \end{bmatrix} \\ &= \tilde{\mathbf{W}}_{RF}^H\mathbf{H}\tilde{\mathbf{F}}_{RF} + \mathbf{Z}', \end{aligned} \tag{7}$$

where $\mathbf{Z}'$ is the $N_W \times N_F$ AWGN matrix coupling with $\tilde{\mathbf{W}}_{RF}$. Vectorizing (7) and using rules for Kronecker product, it becomes

$$\begin{aligned} \mathbf{y}_V &= \text{vec}(\mathbf{Y}) \\ &= (\tilde{\mathbf{F}}_{RF}^T \otimes \tilde{\mathbf{W}}_{RF}^H)\,\text{vec}(\mathbf{H}) + \text{vec}(\mathbf{Z}') \\ &= (\tilde{\mathbf{F}}_{RF}^T \otimes \tilde{\mathbf{W}}_{RF}^H)\,\text{vec}(\mathbf{A}_A\mathbf{D}\mathbf{A}_D^H) + \text{vec}(\mathbf{Z}') \\ &= \underbrace{(\tilde{\mathbf{F}}_{RF}^T \otimes \tilde{\mathbf{W}}_{RF}^H)}_{\mathbf{\Phi}}\,(\mathbf{A}_D^* \otimes \mathbf{A}_A)\,\text{vec}(\mathbf{D}) + \text{vec}(\mathbf{Z}') \\ &= \mathbf{\Phi}\,(\mathbf{A}_D^* \otimes \mathbf{A}_A)\,\text{vec}(\mathbf{D}) + \text{vec}(\mathbf{Z}'), \end{aligned} \tag{8}$$

where $\mathbf{y}_V \in \mathbb{C}^{N_F N_W \times 1}$ is the observation vector.

A. Problem Statement

The CE problem in (8) can be formulated as a least squares problem subject to a sparsity constraint as [8]

$$\begin{aligned} (\hat{\mathbf{A}}_D, \hat{\mathbf{A}}_A, \hat{\mathbf{D}}) &= \arg\min_{\underline{\mathbf{A}}_D,\underline{\mathbf{A}}_A,\underline{\mathbf{D}}} \left\|\mathbf{y}_V - \mathbf{\Phi}\,(\underline{\mathbf{A}}_D^* \otimes \underline{\mathbf{A}}_A)\,\text{vec}(\underline{\mathbf{D}})\right\|_2, \\ \text{s.t.} &\begin{cases} \underline{\mathbf{a}}_D(p) \in \{\tilde{\mathbf{a}}_D(i_D), i_D = 1,\cdots,N_D\}, \\ \underline{\mathbf{a}}_A(p) \in \{\tilde{\mathbf{a}}_A(i_A), i_A = 1,\cdots,N_A\}, \\ \|\text{vec}(\underline{\mathbf{D}})\|_0 = N_P, \end{cases} \end{aligned} \tag{9}$$

where $\tilde{\mathbf{a}}_D(i_D)$ is the i_D^{th} column of $\tilde{\mathbf{A}}_D \in \mathbb{C}^{N_T \times N_D}$, which is the matrix consisting of N_D candidates of the actual departure array propagation vectors presented in a given realization

$$\tilde{\mathbf{A}}_D = \frac{1}{\sqrt{N_T}} \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & e^{j2\pi\frac{1}{N_D}} & \cdots & e^{j2\pi\frac{N_D-1}{N_D}} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & e^{j2\pi\frac{N_T-1}{N_D}} & \cdots & e^{j2\pi\frac{(N_D-1)(N_T-1)}{N_D}} \end{bmatrix}. \tag{10}$$

$\tilde{\mathbf{a}}_A(i_A)$ is the i_A^{th} column of $\tilde{\mathbf{A}}_A \in \mathbb{C}^{N_R \times N_A}$, which is the matrix consisting of N_A candidates of the actual arrival array propagation vectors and has similar structure as the matrix shown in (10). $\underline{\mathbf{D}} \in \mathbb{C}^{N_P \times N_P}$ has N_P non-zero entries. When $N_A = N_R$ and/or $N_D = N_T$, $\tilde{\mathbf{A}}_A$ and/or $\tilde{\mathbf{A}}_D$ are IDFT matrices. (9) can be solved by OMP [8][10].

Once we have the reconstructed channel $\hat{\mathbf{H}} = \hat{\mathbf{A}}_A\hat{\mathbf{D}}\hat{\mathbf{A}}_D^H$, the SVD of the channel can be calculated as

$$\hat{\mathbf{H}} \overset{\text{SVD}}{=} \mathbf{U}\mathbf{\Lambda}\mathbf{V}^H, \tag{11}$$

where the columns of $\mathbf{V} \in \mathbb{C}^{N_T \times N_P}$ and the columns of $\mathbf{U} \in \mathbb{C}^{N_R \times N_P}$ are the right and the left singular vectors of $\hat{\mathbf{H}}$, and the diagonal entries of $\mathbf{\Lambda} = \text{diag}(\lambda_1,\cdots,\lambda_{N_P}) \in \mathbb{R}^{N_P \times N_P}$, $\lambda_1 > \cdots > \lambda_{N_P} > 0$, are the singular values of $\hat{\mathbf{H}}$. The design criterion that can be used for the precoder reconstruction is to minimize the Frobenius norm of the error between the precoder and the right singular vectors of $\hat{\mathbf{H}}$ [6],

$$\begin{aligned} (\hat{\mathbf{F}}_{RF}, \hat{\mathbf{F}}_{BB}) &= \arg\min_{\mathbf{F}_{RF},\mathbf{F}_{BB}} \left\|\mathbf{V}(1:N_S) - \mathbf{F}_{RF}\mathbf{F}_{BB}\right\|_F, \\ \text{s.t. } &\mathbf{f}_{RF}(n_{rf}) \in \{\tilde{\mathbf{f}}_{RF}(l), l = 1,\cdots,N_F\}, n_{rf} = 1,\cdots,N_{RF}, \\ &\|\mathbf{F}_{RF}\mathbf{f}_{BB}(n_s)\|_2 = \|\mathbf{v}(n_s)\|_2, n_s = 1,\cdots,N_S. \end{aligned} \tag{12}$$

The same design criterion leads to the desired combiner.

If the precoder is designed at the receiver, then after the reconstruction, the receiver has to send the information of $\hat{\mathbf{F}}_{RF}$ and $\hat{\mathbf{F}}_{BB}$ to the transmitter, including N_{RF} codebook indices of $\{\hat{\mathbf{f}}_{RF}(n_{rf})\}$ and the matrix $\hat{\mathbf{F}}_{BB}$ of size $N_{RF} \times N_S$. The other solution is to design the precoder at the transmitter, but the feedback overhead is the matrix $\mathbf{V}(1:N_S)$ of size $N_T \times N_S$, which requires more bits, especially for massive MIMO systems.

The complete procedure of the previously proposed CE and precoder/combiner reconstruction can be divided into four steps:

1) CE by OMP at the receiver [8][10].
2) Calculating SVD of the estimated channel matrix at the receiver.
3) Implementing the precoder and the combiner reconstruction by OMP at the receiver [6][10].
4) The receiver sends the information of the reconstructed precoder to the transmitter.

Considering practical applications, each of these four steps is a huge burden for the receiver. Therefore, the motivation of the proposed method is to reduce the computational complexity by avoiding SVD (in the second step) and to reduce the feedback overhead (in the fourth step).

B. The Proposed Algorithm

To avoid SVD, we use a idea that if we approximate the channel matrix with a set of orthogonal vectors, then the decomposition of the channel matrix into orthogonal vectors should just return the same set of vectors up to phase rotations (see **Lemma 2** in the Appendix). Therefore, if the matrix estimates $\hat{\mathbf{A}}_D$ and $\hat{\mathbf{A}}_A$ satisfy $\hat{\mathbf{A}}_D^H\hat{\mathbf{A}}_D = \mathbf{I}$ and $\hat{\mathbf{A}}_A^H\hat{\mathbf{A}}_A = \mathbf{I}$, the precoder and the combiner can be designed directly based on $\hat{\mathbf{A}}_D$ and $\hat{\mathbf{A}}_A$, and the SVD can be omitted, while the same data rate is achievable. More formally, it is stated as follows:

Theorem 1. *Given $\hat{\mathbf{H}} = \hat{\mathbf{A}}_A \hat{\mathbf{D}} \hat{\mathbf{A}}_D^H$, where $\hat{\mathbf{A}}_A \in \mathbb{C}^{N_R \times N_P}$, $\hat{\mathbf{A}}_D \in \mathbb{C}^{N_T \times N_P}$, and $\hat{\mathbf{D}} = diag(\hat{\alpha}_1, \cdots, \hat{\alpha}_{N_P}) \in \mathbb{C}^{N_P \times N_P}$, $|\hat{\alpha}_1| > \cdots > |\hat{\alpha}_{N_P}| > 0$. Let the SVD of $\hat{\mathbf{H}}$ be $\hat{\mathbf{H}} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^H$, where the columns of $\mathbf{U} \in \mathbb{C}^{N_R \times N_P}$ and $\mathbf{V} \in \mathbb{C}^{N_T \times N_P}$ are, respectively, the left and the right singular vectors of $\hat{\mathbf{H}}$, and $\mathbf{\Lambda} = diag(\lambda_1, \cdots, \lambda_{N_P}) \in \mathbb{R}^{N_P \times N_P}$, $\lambda_1 > \cdots > \lambda_{N_P} > 0$. If $\hat{\mathbf{A}}_D^H \hat{\mathbf{A}}_D = \mathbf{I}$ and $\hat{\mathbf{A}}_A^H \hat{\mathbf{A}}_A = \mathbf{I}$, then the data rate $C(\hat{\mathbf{F}}_{RF}, \hat{\mathbf{F}}_{BB}, \hat{\mathbf{W}}_{RF}, \hat{\mathbf{W}}_{BB}) = C(\bar{\mathbf{F}}_{RF}, \bar{\mathbf{F}}_{BB}, \bar{\mathbf{W}}_{RF}, \bar{\mathbf{W}}_{BB})$, where*

1. $(\hat{\mathbf{F}}_{RF}, \hat{\mathbf{F}}_{BB})$ and $(\hat{\mathbf{W}}_{RF}, \hat{\mathbf{W}}_{BB})$ are reconstructed based on $\hat{\mathbf{A}}_D$ and $\hat{\mathbf{A}}_A$, respectively.

2. $(\bar{\mathbf{F}}_{RF}, \bar{\mathbf{F}}_{BB})$ and $(\bar{\mathbf{W}}_{RF}, \bar{\mathbf{W}}_{BB})$ are reconstructed based on $\mathbf{V}$ and $\mathbf{U}$, respectively.

Proof: See the Appendix. ∎

To ensure that $\hat{\mathbf{A}}_D^H \hat{\mathbf{A}}_D = \mathbf{I}$ and $\hat{\mathbf{A}}_A^H \hat{\mathbf{A}}_A = \mathbf{I}$ hold, first, the matrices $\tilde{\mathbf{A}}_D$ and $\tilde{\mathbf{A}}_A$ should be made up of orthogonal vectors. To be formal, one defines the orthogonality of any two vectors as

$$\frac{\langle \tilde{\mathbf{a}}_A(i), \tilde{\mathbf{a}}_A(j) \rangle}{||\tilde{\mathbf{a}}_A(i)||_2 \cdot ||\tilde{\mathbf{a}}_A(j)||_2} = \begin{cases} 0, & i \neq j \\ 1, & i = j \end{cases}, \tag{13}$$

where $\langle \tilde{\mathbf{a}}_A(i), \tilde{\mathbf{a}}_A(j) \rangle$ denotes an inner product of the two vectors. Second, two constraints $\text{rank}(\underline{\mathbf{A}}_D) = N_P$ and $\text{rank}(\underline{\mathbf{A}}_A) = N_P$ are required in order to avoid the repeated vectors in $\hat{\mathbf{A}}_D$ and/or $\hat{\mathbf{A}}_A$. In (9), the estimated matrices $\hat{\mathbf{A}}_D$ and/or $\hat{\mathbf{A}}_A$ could consist of repeated vectors so that $\text{rank}(\hat{\mathbf{A}}_D) < N_P$ and/or $\text{rank}(\hat{\mathbf{A}}_A) < N_P$, which also means that $\hat{\mathbf{A}}_D^H \hat{\mathbf{A}}_D \neq \mathbf{I}$ and/or $\hat{\mathbf{A}}_A^H \hat{\mathbf{A}}_A \neq \mathbf{I}$. In other words, we do not allow that a vector is selected twice from the same codebook.

In addition, as the precoder and the combiner can be reconstructed directly based on $\hat{\mathbf{A}}_D$ and $\hat{\mathbf{A}}_A$, and the column rank of the precoder ($\mathbf{F}_{RF}\mathbf{F}_{BB}$) and the combiner ($\mathbf{W}_{RF}\mathbf{W}_{BB}$) is N_S, only the first N_S (rather than N_P) vectors of $\hat{\mathbf{A}}_D$ and $\hat{\mathbf{A}}_A$ are needed. Consequently, the two constraints can be modified as $\text{rank}(\underline{\mathbf{A}}_D) = N_S$ and $\text{rank}(\underline{\mathbf{A}}_A) = N_S$, and the l_0-norm optimization problem of the CE can be reformulated as

$$(\hat{\mathbf{A}}_D, \hat{\mathbf{A}}_A) = \arg \min_{\underline{\mathbf{A}}_D, \underline{\mathbf{A}}_A} \|\mathbf{y}_V - \mathbf{\Phi}\,(\underline{\mathbf{A}}_D^* \otimes \underline{\mathbf{A}}_A)\,\text{vec}(\underline{\mathbf{D}})\|_2\,,$$
$$\text{s.t.} \begin{cases} \underline{\mathbf{a}}_D(n_s) \in \{\tilde{\mathbf{a}}_D(i_D), i_D = 1, \cdots, N_D\}, \\ \underline{\mathbf{a}}_A(n_s) \in \{\tilde{\mathbf{a}}_A(i_A), i_A = 1, \cdots, N_A\}, \\ \|\text{vec}(\underline{\mathbf{D}})\|_0 = N_S, \\ \text{rank}(\underline{\mathbf{A}}_D) = N_S, \\ \text{rank}(\underline{\mathbf{A}}_A) = N_S. \end{cases} \tag{14}$$

(14) can be solved by OMP [10], which is given in **Algorithm 1**, **Part I**. In **Algorithm 1** Step 7, the selected array propagation vectors will not be selected again in order to ensure that there are no repeated vectors in $\hat{\mathbf{A}}_D$ and $\hat{\mathbf{A}}_A$.

Algorithm 1: CE plus precoder/combiner reconstruction

Input: $\mathbf{y}_V$, $\mathbf{\Phi}$, $\tilde{\mathbf{A}}_D$, $\tilde{\mathbf{A}}_A$, $\tilde{\mathbf{F}}_{RF}$, $\tilde{\mathbf{W}}_{RF}$

Output: $\hat{\mathbf{F}}_{RF}$, $\hat{\mathbf{F}}_{BB}$, $\hat{\mathbf{W}}_{RF}$, $\hat{\mathbf{W}}_{BB}$

1. % **Part I** — CE by OMP
2. $\hat{\mathbf{A}}_D =$ empty matrix, $\hat{\mathbf{A}}_A =$ empty matrix, $\hat{\mathbf{\Psi}} =$ empty matrix, $\mathcal{I}_\mathcal{D} = \varnothing$, $\mathcal{I}_\mathcal{A} = \varnothing$
3. $\mathbf{y}_{tmp} = \mathbf{y}_V$
4. $\mathbf{\Psi} = \mathbf{\Phi}(\tilde{\mathbf{A}}_D^* \otimes \tilde{\mathbf{A}}_A)$
5. for $n_s = 1 : N_S$
6. $\quad \mathbf{g} = \mathbf{\Psi}^H \mathbf{y}_{tmp}$
7. $\quad (\hat{i}_D, \hat{i}_A) = \arg \max_{\substack{i_D \in \{1,\cdots,N_D\} \backslash \mathcal{I}_\mathcal{D} \\ i_A \in \{1,\cdots,N_A\} \backslash \mathcal{I}_\mathcal{A}}} (\mathbf{g}\mathbf{g}^H)_{i,i}$, where $i = (i_D - 1) \cdot N_A + i_A$
8. $\quad \hat{\mathbf{A}}_D = [\hat{\mathbf{A}}_D \,|\, \tilde{\mathbf{a}}_D(\hat{i}_D)]$ and $\hat{\mathbf{A}}_A = [\hat{\mathbf{A}}_A \,|\, \tilde{\mathbf{a}}_A(\hat{i}_A)]$
9. $\quad \hat{\mathbf{\Psi}} = [\hat{\mathbf{\Psi}} \,|\, \mathbf{\Phi} \cdot (\tilde{\mathbf{a}}_D(\hat{i}_D)^* \otimes \tilde{\mathbf{a}}_A(\hat{i}_A))]$
10. $\quad \mathbf{y}_{tmp} = (\mathbf{I}_{N_F N_W} - \hat{\mathbf{\Psi}}(\hat{\mathbf{\Psi}}^H \hat{\mathbf{\Psi}})^{-1} \hat{\mathbf{\Psi}}^H)\mathbf{y}_V$
11. $\quad \mathcal{I}_\mathcal{D} = \mathcal{I}_\mathcal{D} \cup \{\hat{i}_D\}$ and $\mathcal{I}_\mathcal{A} = \mathcal{I}_\mathcal{A} \cup \{\hat{i}_A\}$
12. end
13. $\text{diag}(\hat{\mathbf{D}}) = (\hat{\mathbf{\Psi}}^H \hat{\mathbf{\Psi}})^{-1} \hat{\mathbf{\Psi}}^H \mathbf{y}_V$
14. % Output: $\hat{\mathbf{A}}_D$, $\hat{\mathbf{A}}_A$
15.
16. % **Part II** — The receiver sends the codebook indices of
17. % $\{\hat{\mathbf{a}}_D(n_s), n_s = 1, \cdots, N_S\}$ to the transmitter
18.
19. % **Part III** — Precoder/combiner reconstruction by OMP
20. $\hat{\mathbf{F}}_{RF} =$ empty matrix, $\hat{\mathbf{F}}_{BB} =$ empty matrix
21. $\mathbf{A}_{tmp} = \hat{\mathbf{A}}_D$
22. for $n_{rf} = 1 : N_{RF}$
23. $\quad \mathbf{G} = \tilde{\mathbf{F}}_{RF}^H \mathbf{A}_{tmp}$
24. $\quad \hat{l} = \arg \max_{l=1,\cdots,N_F} (\mathbf{G}\mathbf{G}^H)_{l,l}$
25. $\quad \hat{\mathbf{F}}_{RF} = [\hat{\mathbf{F}}_{RF} \mid \tilde{\mathbf{f}}_{RF}(\hat{l})]$
26. $\quad \mathbf{A}_{tmp} = (\mathbf{I}_{N_T} - \hat{\mathbf{F}}_{RF}(\hat{\mathbf{F}}_{RF}^H \hat{\mathbf{F}}_{RF})^{-1} \hat{\mathbf{F}}_{RF}^H)\hat{\mathbf{A}}_D$
27. end
28. $\hat{\mathbf{F}}_{BB} = (\hat{\mathbf{F}}_{RF}^H \hat{\mathbf{F}}_{RF})^{-1} \hat{\mathbf{F}}_{RF}^H \hat{\mathbf{A}}_D$
29. $\hat{\mathbf{f}}_{BB}(n_s) = \frac{\hat{\mathbf{f}}_{BB}(n_s)}{\|\hat{\mathbf{F}}_{RF} \hat{\mathbf{f}}_{BB}(n_s)\|_2}, n_s = 1, \cdots, N_S$
30. % Use the same procedure to reconstruct $\hat{\mathbf{W}}_{RF}$ and $\hat{\mathbf{W}}_{BB}$

Once $\hat{\mathbf{A}}_D$ is available, N_S codebook indices of $\{\hat{\mathbf{a}}_D(n_s)\}$ have to be sent to the transmitter for the precoder reconstruction. Based on $\hat{\mathbf{A}}_D$, the optimization problem of the precoder reconstruction is similar to (12), which can be written as

$$(\hat{\mathbf{F}}_{RF}, \hat{\mathbf{F}}_{BB}) = \arg \min_{\mathbf{F}_{RF}, \mathbf{F}_{BB}} \left\| \hat{\mathbf{A}}_D - \mathbf{F}_{RF}\mathbf{F}_{BB} \right\|_F\,,$$
$$\text{s.t. } \mathbf{f}_{RF}(n_{rf}) \in \{\tilde{\mathbf{f}}_{RF}(l), l = 1, \cdots, N_F\},$$
$$\|\mathbf{F}_{RF}\mathbf{f}_{BB}(n_s)\|_2 = \|\hat{\mathbf{a}}_D(n_s)\|_2\,. \tag{15}$$

Similarly, given $\hat{\mathbf{A}}_A$, the optimization problem of the combiner reconstruction can be formulated as (15). Note that the codebooks $\tilde{\mathbf{F}}_{RF}$ and $\tilde{\mathbf{W}}_{RF}$ used in the proposed algorithm can be made up of either orthogonal or non-orthogonal vectors. From **Theorem 1**, there are no restrictions on $\tilde{\mathbf{W}}_{RF}$ and $\tilde{\mathbf{F}}_{RF}$; only $\hat{\mathbf{A}}_D$ and $\hat{\mathbf{A}}_A$ are required to be orthogonal matrices.

Table I: Four simulated methods of CE plus HBF.

	Note	CE	SVD	HBF	Feedback
Method 1	*Reference*	No (perfect channel state information (CSI))	Yes	See (12) and [6]	-
Method 2	*Reference*	Formulated as (9) and solved by OMP [8]	Yes	See (12) and [6]	Yes, see Table III
Method 3	*Reference*	**Algorithm 1, Part I** with slight modifications, see (16)	Yes (for reference to demonstrate **Theorem 1**)	See (12) and [6]	-
Method 4	*Proposed*	**Algorithm1, Part I**	No	**Algorithm 1, Part III**	Yes, see **Algorithm 1, Part II** and Table III

Table II: Simulation parameters.

Parameter	Value
Carrier frequency [GHz]	60
Training time (T in (6))	512
Number of paths (N_P)	8
Number of Tx antennas (N_T)	32
Number of Rx antennas (N_R)	32
Number of RF chains (N_{RF})	4
Number of data streams (N_S)	2,4
Number of vectors in $\tilde{\mathbf{A}}_D$ (N_D)	32
Number of vectors in $\tilde{\mathbf{A}}_A$ (N_A)	32
Number of vectors in $\tilde{\mathbf{F}}_{RF}$ (N_F)	32
Number of vectors in $\tilde{\mathbf{W}}_{RF}$ (N_W)	32
Path loss (α_p)	$\sum_{p=1}^{N_P} \lvert\alpha_p\rvert^2 = 1, \alpha_p \in \mathbb{C}$
AoD and AoA (ϕ_p)	$\phi_p \sim \mathcal{U}(-\frac{\pi}{2}, \frac{\pi}{2})$

To sum up, the proposed algorithm can be divided into three steps as detailed in **Algorithm 1**:

1) CE by OMP at the receiver.
2) The receiver sends the codebook indices of $\{\hat{\mathbf{a}}_D(n_s)\}$ to the transmitter.
3) Implementing the precoder reconstruction at the transmitter and the combiner reconstruction at the receiver.

IV. Numerical Results

Four methods of CE plus HBF design are listed in Table I, and Table II lists all the simulation parameters. Since $N_D = N_T$ and $N_A = N_R$, $\tilde{\mathbf{A}}_D$ and $\tilde{\mathbf{A}}_A$ are IDFT matrices which satisfy $\tilde{\mathbf{A}}_D^H\tilde{\mathbf{A}}_D = \mathbf{I}$ and $\tilde{\mathbf{A}}_A^H\tilde{\mathbf{A}}_A = \mathbf{I}$. $\tilde{\mathbf{W}}_{RF}$ and $\tilde{\mathbf{F}}_{RF}$ can be either orthogonal or non-orthogonal matrices in the proposed methods, and in our simulations the orthogonal ones are considered. All the methods use the same codebooks, $\tilde{\mathbf{A}}_D$, $\tilde{\mathbf{A}}_A$, $\tilde{\mathbf{F}}_{RF}$, $\tilde{\mathbf{W}}_{RF}$.

A. Achievable Data Rate

Method 3 is simulated for reference to proof that the proposed algorithm *without* SVD (Method 4) shows the same data rate as the one *with* SVD (Method 3) when the conditions $\hat{\mathbf{A}}_D^H\hat{\mathbf{A}}_D = \mathbf{I}$ and $\hat{\mathbf{A}}_A^H\hat{\mathbf{A}}_A = \mathbf{I}$ hold. As introduced in Section III-B, Method 4 only estimates N_S array propagation vectors in CE, whereas Method 3 estimates N_P vectors. The

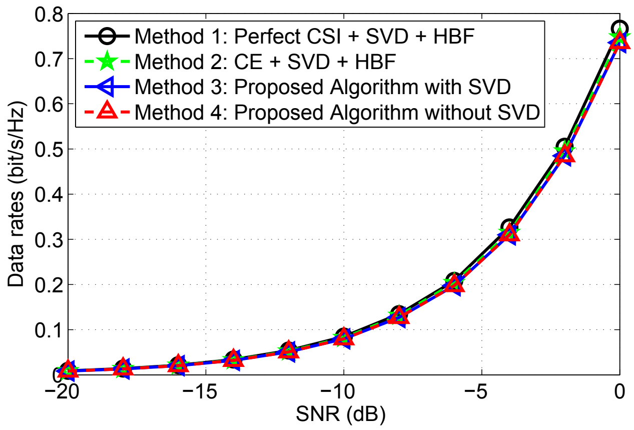

Figure 2: The achievable data rate by various algorithms of CE plus HBF with $N_S = 4$, $N_{RF} = 4$ and $N_P = 8$.

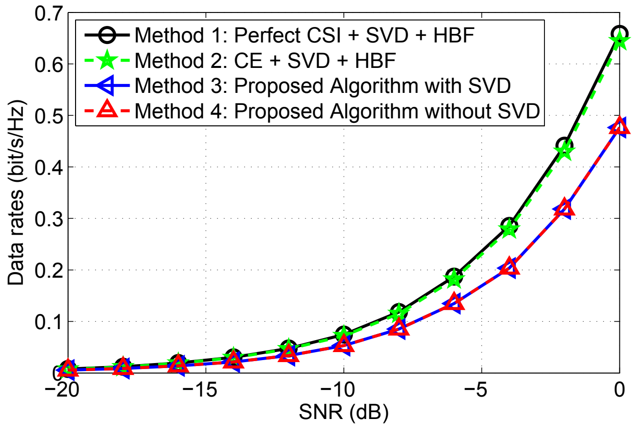

Figure 3: The achievable data rate by various algorithms of CE plus HBF with $N_S = 2$, $N_{RF} = 4$ and $N_P = 8$.

optimization problem of CE in Method 3 is given as follows,

$$(\hat{\mathbf{A}}_D, \hat{\mathbf{A}}_A, \hat{\mathbf{D}}) = \arg \min_{\underline{\mathbf{A}}_D, \underline{\mathbf{A}}_A, \underline{\mathbf{D}}} \|\mathbf{y}_V - \mathbf{\Phi}\left(\underline{\mathbf{A}}_D^* \otimes \underline{\mathbf{A}}_A\right) \mathrm{vec}(\underline{\mathbf{D}})\|_2,$$

$$\text{s.t.} \begin{cases} \underline{\mathbf{a}}_D(p) \in \{\tilde{\mathbf{a}}_D(i_D), i_D = 1, \cdots, N_D\}, \\ \underline{\mathbf{a}}_A(p) \in \{\tilde{\mathbf{a}}_A(i_A), i_A = 1, \cdots, N_A\}, \\ \|\mathrm{vec}(\underline{\mathbf{D}})\|_0 = N_P, \\ \mathrm{rank}(\underline{\mathbf{A}}_D) = N_P, \\ \mathrm{rank}(\underline{\mathbf{A}}_A) = N_P. \end{cases} \quad (16)$$

Table III: Comparison of the feedback overhead by Method 2 and Method 4.

	Feedback overhead	$N_{RF}=4, N_S=4$ $N_F=32, N_D=32$	$N_{RF}=4, N_S=2$ $N_F=32, N_D=32$
Method 2	$N_{RF}\cdot\log_2 N_F + 16\cdot N_{RF}\cdot N_S$	276 bits	148 bits
Method 4	$N_S\cdot\log_2 N_D$	20 bits	10 bits
Reduced feedback overhead	$1-\frac{N_S\cdot\log_2 N_D}{N_{RF}\cdot\log_2 N_F+16\cdot N_{RF}\cdot N_S}$	92.75%	93.24%

The following steps in Method 3 are computing SVD according to the results from (16) and then using the right and the left singular vectors of the estimated channel matrix to reconstruct the precoder and the combiner, respectively.

Fig. 2 and 3 show the achievable data rate by the four simulated methods with different number N_S of data streams. In Fig. 2, since the simulated AoDs and AoAs are uniformly distributed over $(-\frac{\pi}{2},\frac{\pi}{2})$, in Method 2 very few vectors in $\{\hat{\mathbf{a}}_D(p)\}$ and $\{\hat{\mathbf{a}}_A(p)\}$ are repeated; therefore, the resulting data rates of Method 2, 3, and 4 are almost the same. Then, let us see the difference between Fig. 2 and 3. Because the codebooks used in CE and ABF reconstruction are the same, when $N_S = N_{RF}$ (see Fig. 2), the N_{RF} reconstructed ABF vectors are the same as the N_S selected array propagation vectors. Accordingly, the resulting data rates of Method 3 and 4 are desirable. However, when $N_S < N_{RF}$ (see Fig. 3), only N_S (instead of N_{RF}) reconstructed ABF vectors are the same as the N_S selected array propagation vectors, while the others are repeated or invalid (the corresponding weightings in DBF approximate to zero). That is the reason why in Fig. 3 the data rates of Method 3 and 4 degrade significantly; in Method 2, there is no such problem as the HBF reconstruction is based on the right/left singular vectors of the estimated channel matrix.

B. Analyses of the Feedback Overhead

The advantages of the proposed algorithm are not only reducing the computations of SVD but also reducing the feedback overhead. The feedback overhead by Method 2 and 4 are listed in Table III and detailed as follows.

- In Method 2, the transmitter needs the information of
  1) $\hat{\mathbf{F}}_{RF}$: N_{RF} codebook indices of the steering vectors $\{\hat{\mathbf{f}}_{RF}(n_{rf})\}$, which are selected from the set of the N_F vectors defined in $\tilde{\mathbf{F}}_{RF}$. Therefore, $N_{RF}\cdot\log_2 N_F$ bits are required.
  2) $\hat{\mathbf{F}}_{BB}$: DBF matrix of size $N_{RF}\times N_S$. By using 16 bits to represent each entry in $\hat{\mathbf{F}}_{BB}$ including the real and the imaginary parts, $16\cdot N_{RF}\cdot N_S$ bits are required.
- In Method 4, the transmitter needs the information of
  1) $\hat{\mathbf{A}}_D$: N_S codebook indices of the array propagation vectors $\{\hat{\mathbf{a}}_D(n_s)\}$, which are selected from the set of the N_D vectors defined in $\tilde{\mathbf{A}}_D$. Therefore, $N_S\cdot\log_2 N_D$ bits are required.

The reason why the receiver sends the information of $\hat{\mathbf{F}}_{RF}$ and $\hat{\mathbf{F}}_{BB}$ rather than $\mathbf{V}(1:N_S)$ in Method 2 is explained in Section III-A. Comparing these two methods, the proposed one can reduce more than 90% feedback overhead.

V. Conclusions

This paper presents a novel method to reduce the computational complexity and the feedback overhead of the joint CE and HBF problems. By orthogonality of the selected array propagation vectors in the proposed CE method, the precoder and the combiner can be reconstructed *without* SVD of the estimated channel matrix and there is no data rate loss. Also, regarding the feedback overhead, only the codebook indices of the selected array propagation vectors have to be sent to the transmitter. Compared to other joint CE plus HBF reconstruction methods in the literature, the proposed solution can reduce the feedback overhead by more than 90%.

VI. Appendix

Based on the definitions of eigenvalue and eigenvector [14], the following Lemmas show intermediate results of **Theorem 1**.

Lemma 1. *Given $\hat{\mathbf{H}} = \hat{\mathbf{A}}_A\hat{\mathbf{D}}\hat{\mathbf{A}}_D^H$, where $\hat{\mathbf{A}}_A \in \mathbb{C}^{N_R\times N_P}$, $\hat{\mathbf{A}}_D \in \mathbb{C}^{N_T\times N_P}$ and $\hat{\mathbf{D}} = diag(\hat{\alpha}_1,\cdots,\hat{\alpha}_{N_P}) \in \mathbb{C}^{N_P\times N_P}$, where $|\hat{\alpha}_1| > \cdots > |\hat{\alpha}_{N_P}| > 0$. If $\hat{\mathbf{A}}_D^H\hat{\mathbf{A}}_D = \mathbf{I}$ and $\hat{\mathbf{A}}_A^H\hat{\mathbf{A}}_A = \mathbf{I}$, then the columns of $\hat{\mathbf{A}}_D$ and $\hat{\mathbf{A}}_A$ are, respectively, the right and the left singular vectors of $\hat{\mathbf{H}}$.*

Proof: Given a complex diagonal matrix $\hat{\mathbf{D}}$, it can be represented as multiplication of one complex diagonal matrix and one real diagonal matrix, such as $\hat{\mathbf{D}} = \hat{\mathbf{D}}_C\hat{\mathbf{D}}_R$, where $\hat{\mathbf{D}}_C = \text{diag}(e^{j\angle\hat{\alpha}_1},\cdots,e^{j\angle\hat{\alpha}_{N_P}}) \in \mathbb{C}^{N_P\times N_P}$ ($\angle\hat{\alpha}_p$ returns the phase angle of $\hat{\alpha}_p$), $\hat{\mathbf{D}}_C^H\hat{\mathbf{D}}_C = \mathbf{I}$, and $\hat{\mathbf{D}}_R = \text{diag}(|\hat{\alpha}_1|,\cdots,|\hat{\alpha}_{N_P}|) \in \mathbb{R}^{N_P\times N_P}$. Let $\tilde{\mathbf{H}} \triangleq \hat{\mathbf{H}}\hat{\mathbf{H}}^H = \hat{\mathbf{A}}_A\hat{\mathbf{D}}_R^2\hat{\mathbf{A}}_A^H$, which can be further shown as $\tilde{\mathbf{H}}\hat{\mathbf{A}}_A = \hat{\mathbf{A}}_A\hat{\mathbf{D}}_R^2$. From the definition of eigenvector, we know that $\hat{\mathbf{a}}_A(p), p = 1,\cdots,N_P$, are the eigenvectors of $\tilde{\mathbf{H}}$ or the left singular vectors of $\hat{\mathbf{H}}$. Similarly, $\hat{\mathbf{a}}_D(p), p = 1,\cdots,N_P$, are the right singular vectors of $\hat{\mathbf{H}}$. ■

Lemma 2. *Given $\hat{\mathbf{H}} = \hat{\mathbf{A}}_A\hat{\mathbf{D}}\hat{\mathbf{A}}_D^H$ as defined in **Lemma 1**. Let the SVD of $\hat{\mathbf{H}}$ be $\hat{\mathbf{H}} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^H$, where the columns of $\mathbf{U} \in \mathbb{C}^{N_R\times N_P}$ and $\mathbf{V} \in \mathbb{C}^{N_T\times N_P}$ are, respectively, the left and the right singular vectors of $\hat{\mathbf{H}}$, and $\mathbf{\Lambda} = diag(\lambda_1,\cdots,\lambda_{N_P}) \in \mathbb{R}^{N_P\times N_P}$, $\lambda_1 > \cdots > \lambda_{N_P} > 0$. If $\hat{\mathbf{A}}_D^H\hat{\mathbf{A}}_D = \mathbf{I}$ and $\hat{\mathbf{A}}_A^H\hat{\mathbf{A}}_A = \mathbf{I}$, then*
1. $\mathbf{V} = \hat{\mathbf{A}}_D\hat{\mathbf{D}}_D$, where $\hat{\mathbf{D}}_D = diag(e^{j\theta_{D,1}},\cdots,e^{j\theta_{D,N_P}}) \in \mathbb{C}^{N_P\times N_P}$, $0 \le \theta_{D,p} < 2\pi$.
2. $\mathbf{U} = \hat{\mathbf{A}}_A\hat{\mathbf{D}}_A$, where $\hat{\mathbf{D}}_A = diag(e^{j\theta_{A,1}},\cdots,e^{j\theta_{A,N_P}}) \in \mathbb{C}^{N_P\times N_P}$, $0 \le \theta_{A,p} < 2\pi$.

Proof: Because all the eigenvalues are assumed distinct and sorted in descending order. From **Lemma 1**, we know that both $\hat{\mathbf{a}}_A(p)$ and $\mathbf{u}(p)$ are the eigenvectors of $\tilde{\mathbf{H}}$ ($\tilde{\mathbf{H}} \triangleq \hat{\mathbf{H}}\hat{\mathbf{H}}^H$)

corresponding to the same eigenvalue. In addition, eigenvectors corresponding to the same eigenvalue are equivalent up to a phase rotation. Therefore, the relationship between $\hat{\mathbf{a}}_A(p)$ and $\mathbf{u}(p)$ can be written as $\mathbf{u}(p) = e^{j\theta_{A,p}}\hat{\mathbf{a}}_A(p)$, $0 \le \theta_{A,p} < 2\pi$. Similarly, $\mathbf{v}(p) = e^{j\theta_{D,p}}\hat{\mathbf{a}}_D(p)$, $0 \le \theta_{D,p} < 2\pi$. ■

Lemma 3. *Given $\hat{\mathbf{H}} = \hat{\mathbf{A}}_A\hat{\mathbf{D}}\hat{\mathbf{A}}_D^H$ as defined in **Lemma 1**. $(\hat{\mathbf{F}}_{RF}, \hat{\mathbf{F}}_{BB})$ and $(\hat{\mathbf{W}}_{RF}, \hat{\mathbf{W}}_{BB})$ are reconstructed by $\hat{\mathbf{A}}_D$ and $\hat{\mathbf{A}}_A$ respectively. Let the SVD of $\hat{\mathbf{H}}$ be $\hat{\mathbf{H}} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^H$ as defined in **Lemma 2**. $(\bar{\mathbf{F}}_{RF}, \bar{\mathbf{F}}_{BB})$ and $(\bar{\mathbf{W}}_{RF}, \bar{\mathbf{W}}_{BB})$ are reconstructed by $\mathbf{V}$ and $\mathbf{U}$ respectively. If $\hat{\mathbf{A}}_D^H\hat{\mathbf{A}}_D = \mathbf{I}$ and $\hat{\mathbf{A}}_A^H\hat{\mathbf{A}}_A = \mathbf{I}$, then*

*1. $\bar{\mathbf{F}}_{RF} = \hat{\mathbf{F}}_{RF}$ and $\bar{\mathbf{F}}_{BB} = \hat{\mathbf{F}}_{BB}\hat{\mathbf{D}}_D$, where $\hat{\mathbf{D}}_D$ is defined in **Lemma 2**.*

*2. $\bar{\mathbf{W}}_{RF} = \hat{\mathbf{W}}_{RF}$ and $\bar{\mathbf{W}}_{BB} = \hat{\mathbf{W}}_{BB}\hat{\mathbf{D}}_A$, where $\hat{\mathbf{D}}_A$ is defined in **Lemma 2**.*

Proof: To proof $\bar{\mathbf{F}}_{RF} = \hat{\mathbf{F}}_{RF}$ is equivalent to proofing that the results of $\mathbf{G}\mathbf{G}^H$ (in **Algorithm 1** Step 24) for the two different observations, $\mathbf{V}$ and $\hat{\mathbf{A}}_D$, in every iteration are the same. Also, since $\mathbf{G}\mathbf{G}^H = \tilde{\mathbf{F}}_{RF}^H\mathbf{A}_{tmp}\mathbf{A}_{tmp}^H\tilde{\mathbf{F}}_{RF}$ and $\tilde{\mathbf{F}}_{RF}$ is the given codebook, the problem can be further simplified as proofing that the results of $\mathbf{A}_{tmp}\mathbf{A}_{tmp}^H$ for the two different observations are the same in every iteration.

When $n_{rf} = 1$,

$$\begin{aligned}\bar{\mathbf{A}}_{tmp}\bar{\mathbf{A}}_{tmp}^H &= \mathbf{V}\mathbf{V}^H \\ &\overset{\text{Lemma 2}}{=} \hat{\mathbf{A}}_D\hat{\mathbf{D}}_D\hat{\mathbf{D}}_D^H\hat{\mathbf{A}}_D^H \\ &= \hat{\mathbf{A}}_D\hat{\mathbf{A}}_D^H \\ &= \hat{\mathbf{A}}_{tmp}\hat{\mathbf{A}}_{tmp}^H. \end{aligned} \tag{17}$$

(17) ensures that the first selected ABF vectors by $\mathbf{V}$ and $\hat{\mathbf{A}}_D$ are the same. Then, the updated $\hat{\mathbf{A}}_{tmp}$ and $\bar{\mathbf{A}}_{tmp}$ become

$$\hat{\mathbf{A}}_{tmp} = (\mathbf{I}_{N_T} - \hat{\mathbf{F}}_{RF}(\hat{\mathbf{F}}_{RF}^H\hat{\mathbf{F}}_{RF})^{-1}\hat{\mathbf{F}}_{RF}^H)\hat{\mathbf{A}}_D \tag{18}$$

and

$$\begin{aligned}\bar{\mathbf{A}}_{tmp} &= (\mathbf{I}_{N_T} - \bar{\mathbf{F}}_{RF}(\bar{\mathbf{F}}_{RF}^H\bar{\mathbf{F}}_{RF})^{-1}\bar{\mathbf{F}}_{RF}^H)\mathbf{V} \\ &= (\mathbf{I}_{N_T} - \hat{\mathbf{F}}_{RF}(\hat{\mathbf{F}}_{RF}^H\hat{\mathbf{F}}_{RF})^{-1}\hat{\mathbf{F}}_{RF}^H)\hat{\mathbf{A}}_D\hat{\mathbf{D}}_D \\ &= \hat{\mathbf{A}}_{tmp}\hat{\mathbf{D}}_D. \end{aligned} \tag{19}$$

When $n_{rf} > 1$, according to the updated $\bar{\mathbf{A}}_{tmp}$ and $\hat{\mathbf{A}}_{tmp}$, one has

$$\begin{aligned}\bar{\mathbf{A}}_{tmp}\bar{\mathbf{A}}_{tmp}^H &= \hat{\mathbf{A}}_{tmp}\hat{\mathbf{D}}_D\hat{\mathbf{D}}_D^H\hat{\mathbf{A}}_{tmp}^H \\ &= \hat{\mathbf{A}}_{tmp}\hat{\mathbf{A}}_{tmp}^H. \end{aligned} \tag{20}$$

Repeat (18)–(20) until $n_{rf} = N_{RF}$. It shows that the results of $\mathbf{A}_{tmp}\mathbf{A}_{tmp}^H$ for the two different observations are the same in every iteration and, consequently, it can be concluded that the selected ABF vectors are equal, i.e., $\bar{\mathbf{F}}_{RF} = \hat{\mathbf{F}}_{RF}$.

Based on the results that $\bar{\mathbf{F}}_{RF} = \hat{\mathbf{F}}_{RF}$ and $\mathbf{V} = \hat{\mathbf{A}}_D\hat{\mathbf{D}}_D$ (from **Lemma 2**), the DBF matrix by the observation $\mathbf{V}$ is shown as (see **Algorithm 1** Step 28)

$$\begin{aligned}\bar{\mathbf{F}}_{BB} &= (\bar{\mathbf{F}}_{RF}^H\bar{\mathbf{F}}_{RF})^{-1}\bar{\mathbf{F}}_{RF}^H\mathbf{V} \\ &= (\hat{\mathbf{F}}_{RF}^H\hat{\mathbf{F}}_{RF})^{-1}\hat{\mathbf{F}}_{RF}^H\hat{\mathbf{A}}_D\hat{\mathbf{D}}_D \\ &= \hat{\mathbf{F}}_{BB}\hat{\mathbf{D}}_D. \end{aligned}$$

Similarly, we have $\bar{\mathbf{W}}_{RF} = \hat{\mathbf{W}}_{RF}$ and $\bar{\mathbf{W}}_{BB} = \hat{\mathbf{W}}_{BB}\hat{\mathbf{D}}_A$. ■

Theorem 1. *Given $\hat{\mathbf{H}} = \hat{\mathbf{A}}_A\hat{\mathbf{D}}\hat{\mathbf{A}}_D^H$ as defined in **Lemma 1**. Let the SVD of $\hat{\mathbf{H}}$ be $\hat{\mathbf{H}} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^H$ as defined in **Lemma 2**. If $\hat{\mathbf{A}}_D^H\hat{\mathbf{A}}_D = \mathbf{I}$ and $\hat{\mathbf{A}}_A^H\hat{\mathbf{A}}_A = \mathbf{I}$, then the data rate $C(\hat{\mathbf{F}}_{RF}, \hat{\mathbf{F}}_{BB}, \hat{\mathbf{W}}_{RF}, \hat{\mathbf{W}}_{BB}) = C(\bar{\mathbf{F}}_{RF}, \bar{\mathbf{F}}_{BB}, \bar{\mathbf{W}}_{RF}, \bar{\mathbf{W}}_{BB})$, where $(\hat{\mathbf{F}}_{RF}, \hat{\mathbf{F}}_{BB})$ and $(\hat{\mathbf{W}}_{RF}, \hat{\mathbf{W}}_{BB})$ are respectively reconstructed by $\hat{\mathbf{A}}_D$ and $\hat{\mathbf{A}}_A$, and $(\bar{\mathbf{F}}_{RF}, \bar{\mathbf{F}}_{BB})$ and $(\bar{\mathbf{W}}_{RF}, \bar{\mathbf{W}}_{BB})$ are respectively reconstructed by $\mathbf{V}$ and $\mathbf{U}$.*

Proof: By using the relationship between $(\hat{\mathbf{F}}_{RF}, \hat{\mathbf{F}}_{BB}, \hat{\mathbf{W}}_{RF}, \hat{\mathbf{W}}_{BB})$ and $(\bar{\mathbf{F}}_{RF}, \bar{\mathbf{F}}_{BB}, \bar{\mathbf{W}}_{RF}, \bar{\mathbf{W}}_{BB})$ in **Lemma 3**, the data rates (5) by these two HBF designs show the same results. ■

ACKNOWLEDGMENT

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

REFERENCES

[1] T. S. Rappaport, et al., *Wireless Communications: Principles and Practice*, 2nd Edition, Pearson Education, Upper Saddle River, NJ, 2002.
[2] T. A. Thomas, et al., "3D mmWave Channel Model Proposal," *IEEE VTC Fall*, pp. 1–6, 2014.
[3] T. S. Rappaport, et al., "Millimeter Wave Mobile Communications for 5G Cellular: It Will Work!," *IEEE Access*, Vol. 1, pp. 335–349, May 2013.
[4] A. Hajimiri, et al., "Integrated Phased Array Systems in Silicon," *Proceedings of the IEEE*, Vol. 93, pp. 1637–1655, Sep. 2005
[5] J. C. Liberti, et al., *Smart Antennas for Wireless Communications: IS-95 and Third Generation CDMA Applications*, Prentice Hall, NJ, 1999.
[6] O. E. Ayach, et al., "Low Complexity Precoding for Large Millimeter Wave MIMO Systems," *IEEE International Conference on Communications*, pp. 3724–3729, Jun. 2012.
[7] A. Alkhateeb, et al, "Channel Estimation and Hybrid Precoding for Millimeter Wave Cellular Systems," *IEEE Journal of Selected Topics in Signal Processing*, Vol.8, pp. 831–846, Jul. 2014.
[8] R. Mendez-Rial, et al, "Channel estimation and hybrid combining for mmWave: Phase shifters or switches?" *Information Theory and Applications Workshop*, pp. 90–97, 2015.
[9] E. Candes and T. Tao, "Decoding by Linear Programming," *IEEE Transactions on Information Theory*, Vol. 51, pp. 4203–4215, Dec. 2005.
[10] 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, Jul. 2011.
[11] S. Foucart and H. Rauhut, *A Mathematical Introduction to Compressive Sensing*, Birkhauser, 2013.
[12] V. Frascolla, et al., "Challenges and Opportunities for Millimeter-wave Mobile Access Standardization" *IEEE Workshop on Globecom*, pp. 553–558, 2014.
[13] A. Goldsmith, S. 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, 2003.
[14] C. D. Meyer, *Matrix Analysis and Applied Linear Algebra*, 2000.